Towards understanding how young Japanese female college students pronounce the letters of English alphabet - Part III: Analyses of the formants and pitch percentiles on the basis of the linear model

Guido Izuta

『山形県立米沢女子短期大学紀要』 第51号

2015年12月　日刊　別刷

Towards understanding how young Japanese female college students pronounce the letters of English alphabet - Part III: Analyses of the formants and pitch percentiles on the basis of the linear model

Guido Izuta

Abstract

This paper follows up the investigations on how young Japanese female college students read the letters of the English alphabet. In the previous reports, findings related to the analysis of formants F1 and F2 and pitch modulations were presented. Here, a characterization of the percentiles of the formants as well pitch frequencies is carried out on the grounds of the statistical linear model. Yet, in this research, the data of the British English speakers was added and the number of the speakers of the American English was increased to the analyses. The results showed that the students had basically the linear model fitted to the percentiles of both the formants as well pitch frequencies for all the utterances considered whereas the native speakers had also some nonlinear models describing the temporal variations of the sounds.

Keywords

(1) English alphabet pronunciation, (2) English sounds, (3) percentiles of the voice sounds, (4) formant and pitch analyses, (5) linear model of the voice percentiles.

1 INTRODUCTION

To greater or lesser extent, the study of pronunciation has been the focus of attention of both linguistics and phonology (Jack and Theodore, 2001; Ladefoged, 1993; Ladefoged, 2007; Titze 1994). In the language learning settings, it has long been part of the curriculum and a number of approaches have been developed to help the leaners in the process of language assimilation (Morley, 1991). In fact, Moore (2001) showed by means of an experimental and comparative study in Germany that their students of English language tended to improve their communication skills when continuous and direct instructions on how to pronounce the words were given rather than just exposing them to the target language. In the context of English language learning by Japanese learners, there are a lot of reports primarily on case studies related to the difficulties that the students have in

producing the English sounds and learning the language (Riney and Anderson-Hsieh, 1993; Ohata, 2004; Smith, 2012).

In the acoustic analysis of utterances, two elements play the central role in the characterization of the sounds; namely the formants and pitch frequencies. The frequencies of the former correspond to the movement of the tongue and lips during the sound generation (Ladefoged, 2007), so that they allow us to say, to some extent, how two individuals are moving their mouths one relatively to the other. On the other hand, the latter are related to the perception of the sounds (Fujisaki and Kawashima, 1968). For example, the influence of pitch frequencies on the vowel perception was studied by Barreda and Nearey (2012) whereas Binns and Culling (2007) investigated how the external disturbances affect the speech perception, and Carrol et al (2012) looked into the speech perception of cochlear implant users when the reading of sentences was made in the presence of noise. As a matter of fact, the pitch frequencies have been researched from many standpoints. Sugimoto (2014) was concerned with the models to study the perception of accent; Hillenbrand and Clark (2009) were interested in the influence on the voice recognition mechanism that recognizes the male and female voices; and recently many researchers have focused on the neural structure in charge of the pitch information processing (see for example Patterson and Johnstrude (2012), and Zatorre and Gandour (2008))

Motivated by these findings, our research group has examined how young Japanese female students read the English alphabet; and has shown that there are statistical differences in the formants when compared with native speakers of North American English (Izuta, 2013). Furthermore, a correlation testing between the English vowel sounds made by Japanese students and a set of Japanese language sounds similar to some of the English phonemes showed that, nevertheless the Japanese students tried to make the English utterances when they were instructed to do so, the group as a whole generated voice formants statistically similar to the Japanese sounds rather than those of the native English speakers (Izuta, 2014a). Yet, the analysis of the formants F1 and F2 suggested that the students made variant forms of the utterances when compared to the North American English sounds (Izuta, 2014b). The investigation of pitch modulations showed that the groups of students modulated the pitch frequencies differently from the group of native speakers (Izuta, 2015a). Experiments to follow up these results were recently carried out by increasing the number of native speakers. Actually, the data set of the sounds measured out of the American English

speakers was thoroughly renewed with the number of individual being as double as in the previous set, and the data processing was carried out from the very beginning. Moreover, a set of native speakers of British English was added to the study in order to check whether a particular variety of spoken English has any predominant influence on the way the students speak the sounds of the alphabet. The results of the following up experiments aimed to the analyses of the formants and the percentiles were briefly discussed in Izuta (2015b) and Izuta (2015c). In addition, the general analysis of the pitch frequencies was presented in Izuta (2014d) and the percentiles were considered in Izuta (2015e).

In this work, the percentiles of the groups - defined as JP for the group of all the students taking part in the experiment; S1 for the group of 1st year students enrolled in social sciences course; E1 and E2 standing for the groups of 1st year students and 2nd year students from the English language department; and UK and US representing the groups of native speakers of British English and North American English respectively – were characterized by means of the statistical linear model.

Finally, the paper is organized as follows: the experimental protocol is given in section 2; the results in section 3; and the discussions in section 4.

2 EXPERIMENTAL PROCEDURE

The experimental setup was basically the same as reported in our previous papers. In what follows, a brief description of the subjects and data processing procedures are given.

2.1 Subjects

The group of students was composed by female college students from the Tohoku region and aged 18 to 21 without any health problems. The group JP of all students had 26 individual. Splitting the group JP, we have the group S1 of 1st year students enrolled in social sciences course with 9 students, the groups E1 and E2 of 1st and 2nd year students, each with 9 and 8 students. Still, the groups UK and US of native speakers of British English and North American English respectively had 10 and 20 subjects in their 20s and 30s.

2.2 Data acquisition

Before the measuring sessions, the Japanese students participated in training sessions.

During the practicing time, the students made use of whatever audiovisual resources they wanted in order to check for the reading of the letters of the English alphabet. Amongst them, digital dictionaries (Cambridge, 2004; Collins, 2014; forvo, 2014; Weblio, 2014) as well as web resources. Two types of data acquisition sessions were carried out. First, the voice sounds were recorded inside the laboratory in front of a PC; and another carried out outside the school building simulating an interview session. The data set of the native speakers was acquired on the basis of the street interviewing and through the internet.

2.3 Data processing

The data was processed on a Microsoft windows 8 computer equipped with software applications Pratt, excel and R.

3 RESULTS

In this section we present the results for all the letters considered. The graphs have a horizontal axis standing for the formant F2, a vertical axis on the left representing the values of the pitch frequency percentiles, and another vertical axis on the right for the percentiles of the formant F1. The linear fitting curves are also drawn on the figure. Due to the multiple readings that the letter ‘z’ have, the analysis of this letter was not considered here.

3.1 Letter ‘A’

Figures 1 through 6 illustrate the percentiles of the formants F1 as well as pitch frequencies relatively to the formants F2. Expressing the best fit linear curves by the triplet (slope, intercept, r-squared), the group JP yielded (0.8847, -975.03, 0.9123), whereas S1 had (0.9227, -1024.4, 0.8303); E1, (0.8111, -808.09, 0.9369); E2, (0.9073, -1077.80, 0.9137); UK, (0.6662, -743.54, 0.6493); and US, (0.4708, -308.06, 0.8105). Unlike the percentiles of the group of Japanese students, which had r-squared values greater than 0.9, the points of the group of native speakers were best fitted by means of nonlinear curves. Thus, the Japanese students positioned the tongue back and forth proportionally to the open/close movement of the lips. On the other hand, the fitting curves of the percentiles of pitch frequencies were characterized by the triplets (0.0742, 78.784, 0.9781) for the group JP, and (0.093, 33.184, 0.9708) for S1, (0.0489, 144.7,

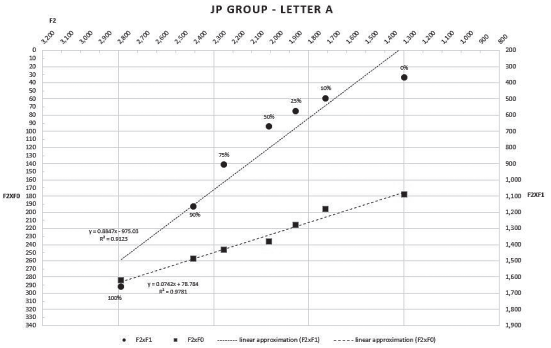

Fig 1. Letter ‘A’ - Group JP.

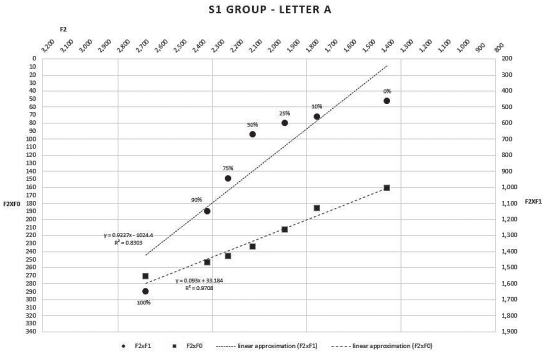

Fig 2. Letter ‘A’ - Group S1

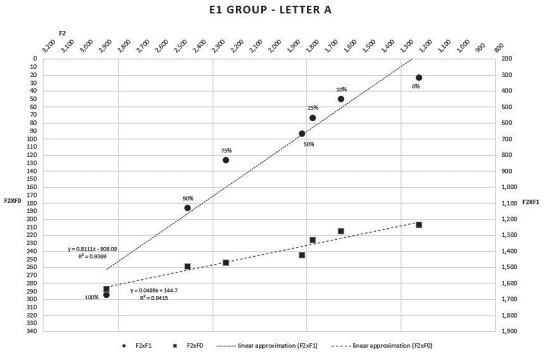

Fig 3. Letter ‘A’ - Group E1

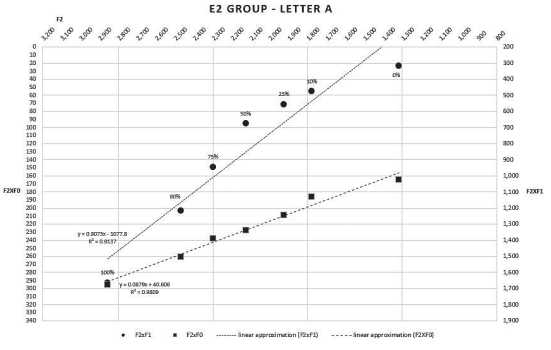

Fig 4. Letter ‘A’ - Group E2

0.9415) for E1, (0.0879, 40.606, 0.9809) for E2, (0.0666, 84.006, 0.6779) for UK, and (0.0625, 82.274, 0.8863) for US. As in the case of the formants, the group of UK and US had nonlinear curves better fitting to the pitch percentiles. Moreover, the slopes of the students' fitting curves were in general steeper than the natives', meaning that the values of the pitch frequencies had linearly varying values with high variation ratios.

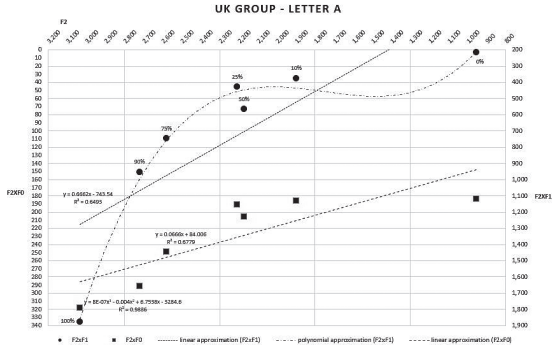

Fig 5. Letter 'A' - Group UK

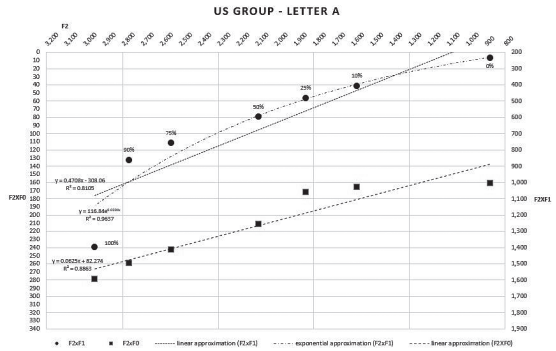

Fig 6. Letter 'A' - Group US

3.2 Letter 'B'

For this letter, the group JP of Japanese students had formants characterized by the trio (1.0759, -1494.2, 0.9253) as shown in Fig.7, and (1.0549, -1451.9, 0.8998) for S1 (Fig.8), (1.1058, -1565.8, 0.971) for E1 (Fig. 9), (1.0388, -1402.1, 0.883) for E2 (Fig. 10), (0.8335, -1298.6, 0.572) for UK (Fig. 11), and (0.5619, -635.72, 0.6156) for US as depicted in Fig. 12. Note that the fitting curves are highly linear for the group consisting of Japanese students, whereas

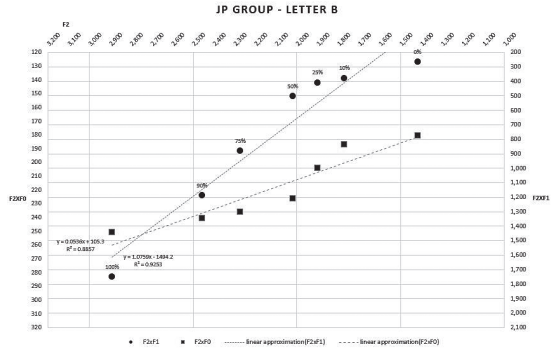

Fig 7. Letter 'B' - Group JP

Fig 8. Letter 'B' - Group S1

the group of native respondents made utterances which were much better expressed by a third order polynomial. This sound behavior along the time line suggests the existence of some kind of frequency modulation.

As for the percentiles of pitch frequencies, the group of Japanese students measured (0.0536, 105.3, 0.8857) for JP; and S1 gave (0.058, 99.188, 0.9375); E1, (0.0407, 144.81, 0.7925); E2, (0.0605, 72.288, 0.8724). For the group of native speakers, the group UK yielded (0.0671, 65.536, 0.8279) whereas the set (0.0629, 63.651, 0.9628) was obtained for the group US. Looking at the r-squared values, they show that all the curves were fairly well described by linear curves. Comparing the group of Japanese students with the natives, the slopes indicate that only E2 had a fitting line similar to the linear curves measured for the natives.

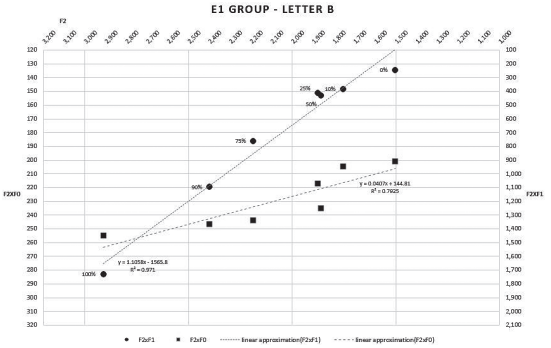

Fig 9. Letter 'B' - Group E1

Fig 10. Letter 'B' - Group E2

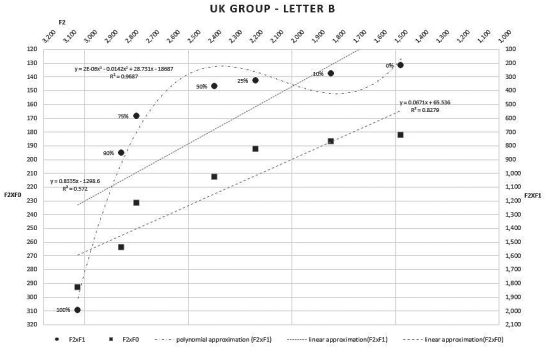

Fig 11. Letter 'B' - Group UK

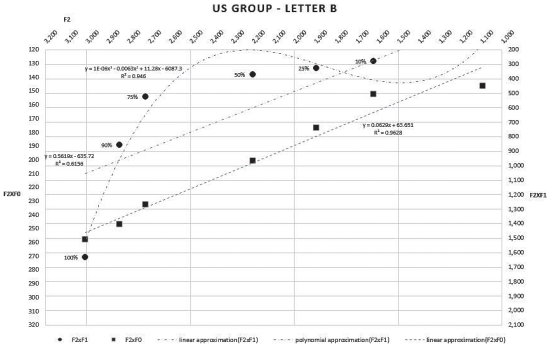

Fig 12. Letter 'B' - Group US

3.3 Letter ‘C’

The values of r-squared for the letter ‘c’ show that both the students and the natives subjects had the percentiles approximated by linear curves; in fact, these curves read (1.1814, -1657.4, 0.9256) for the group JP, whereas S1 probed (1.2354, -1769.1, 0.9049); and E1, (1.0854, -1424.4, 0.9238); E2, (1.2307, -1805.8, 0.9408); UK, (1.0092, -1567.8, 0.8413); US, (0.8023, -1017.9, 0.8707). Nevertheless, the percentiles of the native speakers were much highly correlated to a second order polynomial for UK, and exponential curve for US. These mean that it is reasonable to assume that the native speakers were moving their tongues and lips in order to modulate the sound frequencies during their generation.

As far as the percentiles of the pitch frequencies are concerned, Fig.13 displays that the group JP had a fitting curve expressed by (0.1028, 19.032, 0.9523), and figures 14 through 18 give (0.1046, 16.597, 0.9482) for S1,

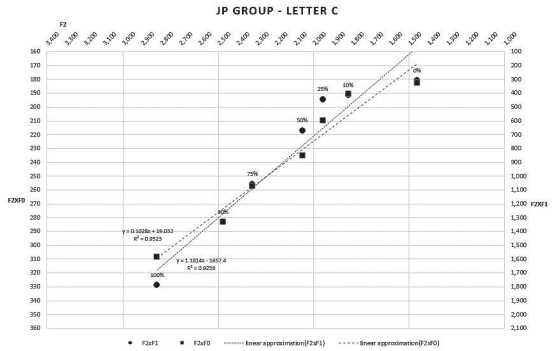

Fig 13. Letter ‘C’ – Group JP

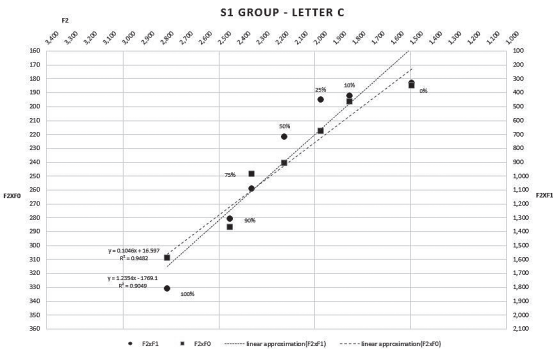

Fig 14. Letter ‘C’ – Group S1

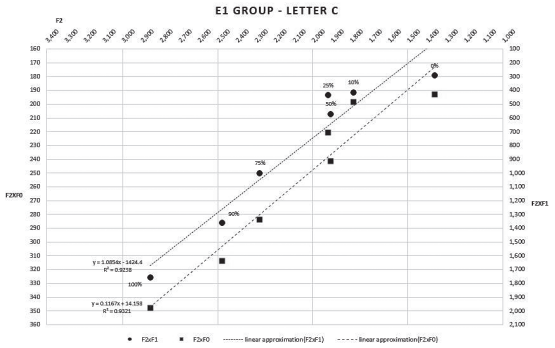

Fig 15. Letter ‘C’ – Group E1

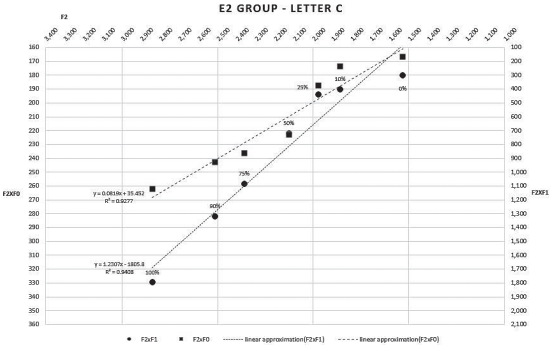

Fig 16. Letter ‘C’ – Group E2

(0.1167, 14.158, 0.9321) for E1, (0.0819, 35.452, 0.9277) for E2, (0.0868, 28.412, 0.7937) for UK, and (0.0654, 60.973, 0.9251) for US. The r-squared values indicate that ruling out the UK group, the curves are highly linear. Yet, the slopes of the curves were relatively a bit smaller for the natives. However, this fact does not mean that they were statistically different; in fact, here, the focus is only on the temporal behavior.

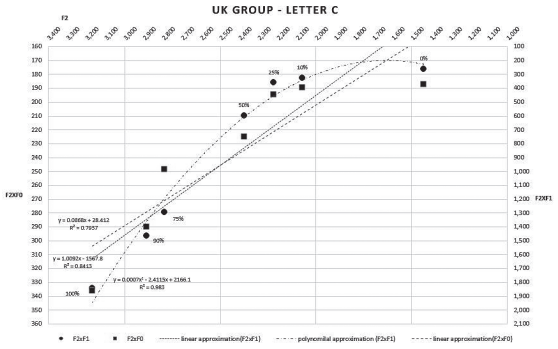

Fig 17. Letter 'C' – Group UK

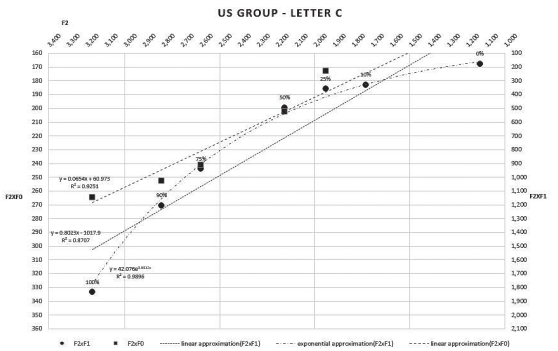

Fig 18. Letter 'C' – Group US

3.4 Letter 'D'

As seen in figures 19 through 24, the percentiles of the formants had fitting curves which were relatively highly linear for the Japanese students and nonlinearly behaving curves - namely, polynomials – for the groups of natives. Indeed, Fig. 19 shows the triplet (1.0305, -1383.2, 0.9115) for JP, whereas the remaining plots led to the trio with values as (1.1108, -1541.6, 0.874) for S1, (0.9963, -1293.7, 0.932) for E1, (0.971, -1290.3, 0.912) for E2, (1.0643,

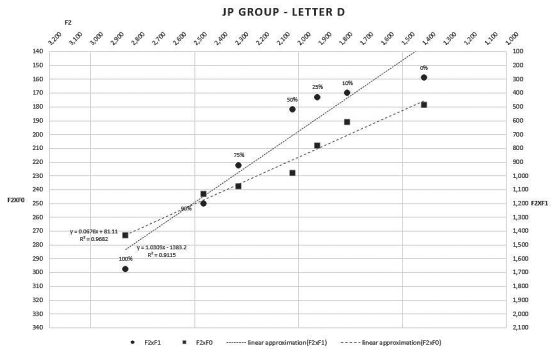

Fig 19. Letter 'D' – Group JP

Fig 20. Letter 'D' – Group S1

-1845.5, 0.699) for UK, and (0.6422, -783.73, 0.658) for US. These nonlinearities in the sound productions by the natives reflect mostly the forward/backward positioning of the tongue during the first half of the voice generation process.

As for the percentiles of the pitch frequencies, the rounded values of r-squared were all greater than or equal to 0.9, so that it is reasonable to consider that they were all linearly varying curves. Focusing on figures 19 through 24, it is clear that the group JP yielded (0.0676, 81.11, 0.9682); S1, (0.0862, 48.549, 0.9294); E1, (0.0553, 117.12, 0.9223); E2, (0.0609, 75.953, 0.9168); UK, (0.0747, 40.172, 0.9249); US, (0.0715, 45.63, 0.8996). For the group of the students, the slope of S1 was a bit larger than the other groups. On the other hand, the groups JP, E1, and E2 were relatively smaller than the values measured for the native speakers.

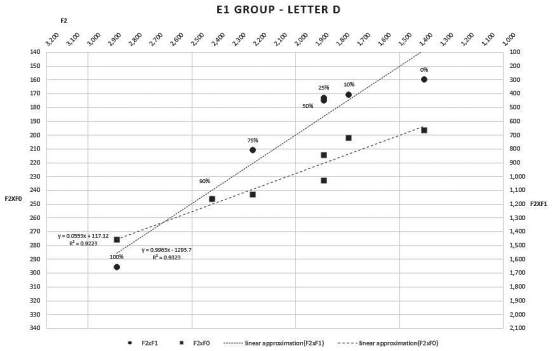

Fig 21. Letter ‘D’ – Group E1

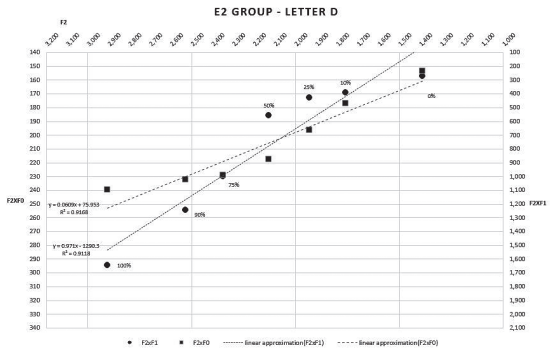

Fig 22. Letter ‘D’ – Group E2

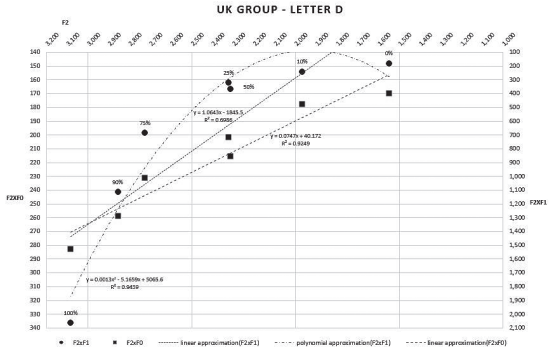

Fig 23. Letter ‘D’ – Group UK

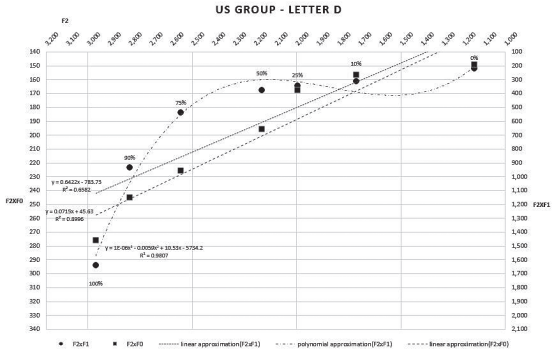

Fig 24. Letter ‘D’ – Group US

3.5 Letter 'E'

The data analysis of the sounds of this letter rendered the triplet (0.9557, -1268.3, 0.8825) for the group JP (Fig.25), (0.9131, -1172.2, 0.8614) for S1 (Fig. 26), (1.0133, -1377.4, 0.9332) for E1 (Fig. 27), (0.9172, -1205, 0.8317) for E2 (Fig. 28), (0.7291, -1056.5, 0.5259) for UK (Fig. 29), (0.4111, -324.81, 0.621) for US (Fig. 30). Unlike the curves of the native speakers, which had higher correlation with third order degree polynomial rather than lines, the group of Japanese students had percentiles correlating positively with linear curves; i.e. with r-squared values greater than 0.85. Still, following the percentiles of the natives, the plots show that the lips opened slightly to then close with a forward positioning of the tongue, and finally opened widely again.

As for the percentiles of the pitch frequencies, the following values were measured: (0.0734, 66.43, 0.9807) for the group JP; (0.0879, 44.267, 0.9791) for S1, (0.0578, 100.74, 0.8807)

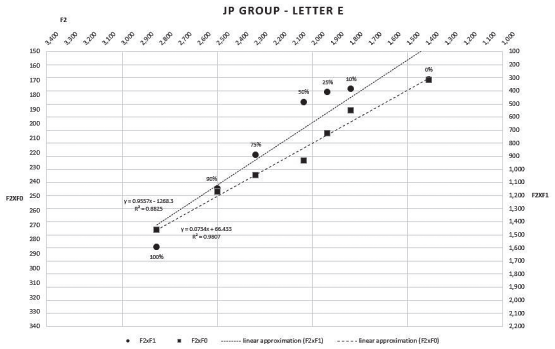

Fig 25. Letter 'E' – Group JP

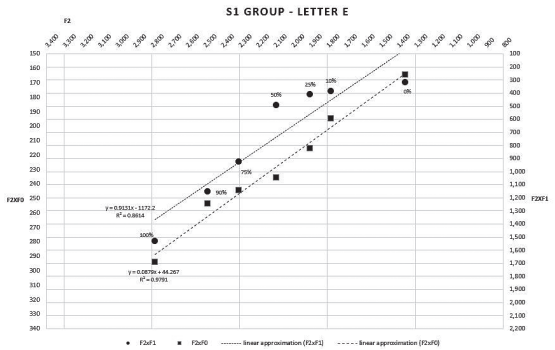

Fig 26. Letter 'E' – Group S1

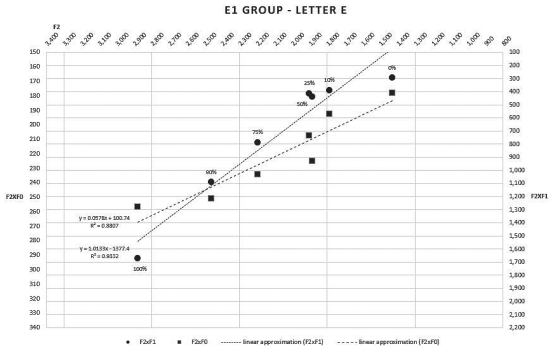

Fig 27. Letter 'E' – Group E1

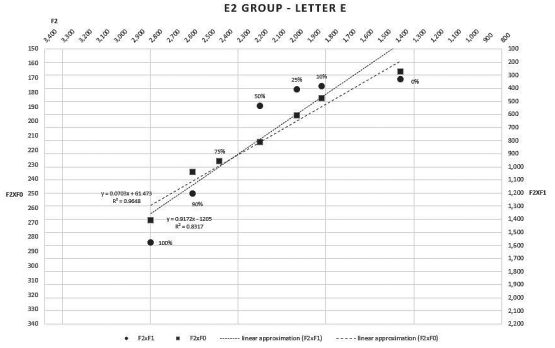

Fig 28. Letter 'E' – Group E2

for E1, (0.0703, 61.473, 0.9648) for E2, (0.038, 130.58, 0.6314) for UK, and (0.0504, 108.12, 0.9419) for US. Ruling out the group UK, the other groups had percentiles well represented by linear curves with r-squared values greater than 0.85. Furthermore, E1 had a slope with r-squared value close to the one of US whereas the other groups of students had steeper variations.

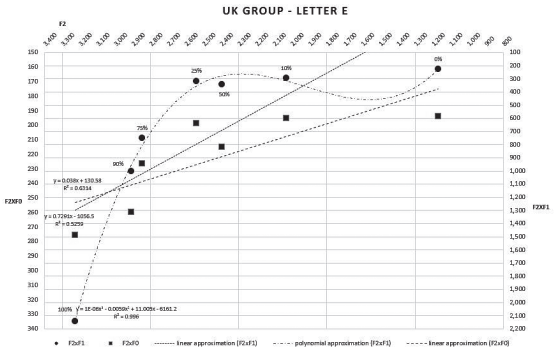

Fig 29. Letter 'E' – Group UK

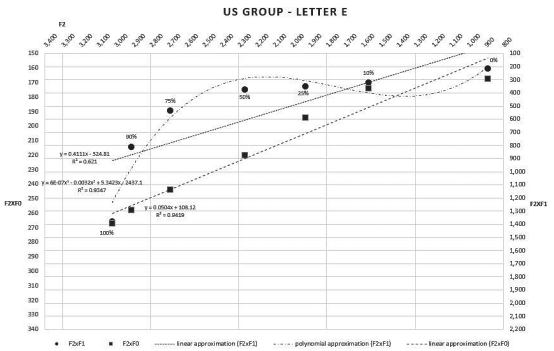

Fig 30. Letter 'E' – Group US

3.6 Letter 'F'

Figures 31 through 36 provide the triplet characterizing the fitting curves of the percentiles of the formants for all the groups analyzed. In fact, for the group JP, the trio was (1.0547, -1121.3, 0.9349), and for S1 it was found the set (1.189, -1408.8, 0.9166), for E1 it read (0.9482, -894.77, 0.9316), for E2 it translated into (1.0139, -1035.2, 0.9394), for UK it was (1.089, -1222.2, 0.9416), and for the group US the triplet was (0.83, -658.36, 0.985). From these, it is clear that the fitting curves

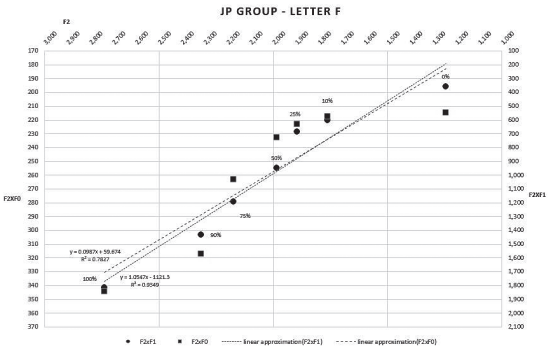

Fig 31. Letter 'F' – Group JP

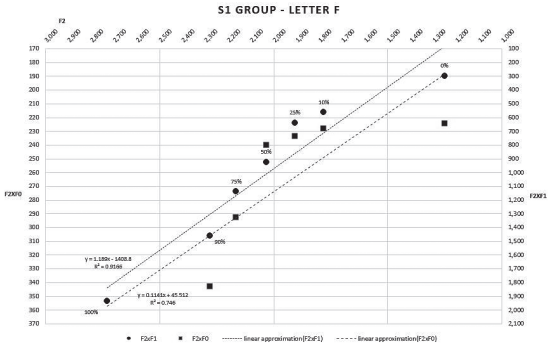

Fig 32. Letter 'F' – Group S1

were well represented by linear lines, meaning that the opening movement of the lips and the positioning of the tongues were correlated in a proportional fashion, so that the movement of one led the other to behave similarly in a linear fashion. Nevertheless a formal statistical testing is necessary to figure out the differences in the slopes, the values here show that the slopes of the students were qualitatively close to the curves computed for the natives.

Now, the percentiles of the pitch frequencies were (0.0987, 59.674, 0.7827) for the group JP, (0.1141, 45.512, 0.746) for S1, (0.0894, 81.732, 0.7975) for E1, (0.0917, 51.266, 0.7645) for E2, (0.0754, 53.872, 0.8537) for UK, and (0.0833, 62.333, 0.8648) for US. From the r-squared values, the curves of the natives were more linearly correlated than those of the students, which suggests that the students were trying to modulate the pitch frequencies along the time scale as they uttered.

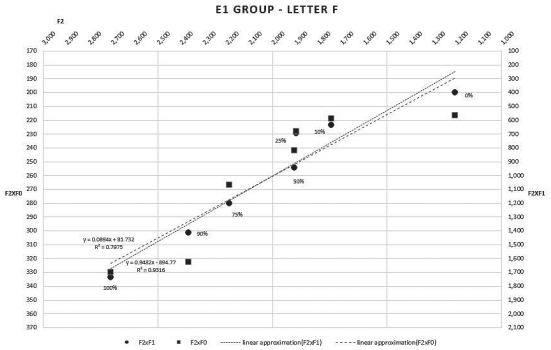

Fig 33. Letter 'F' – Group E1

Fig 34. Letter 'F' – Group E2

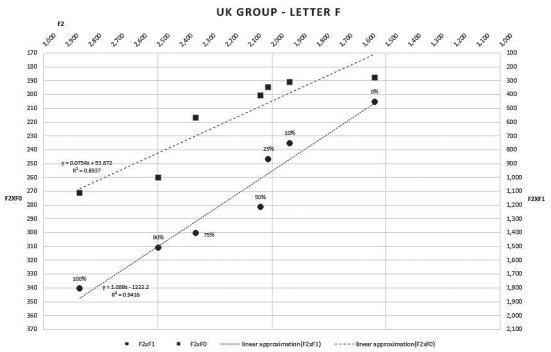

Fig 35. Letter 'F' – Group UK

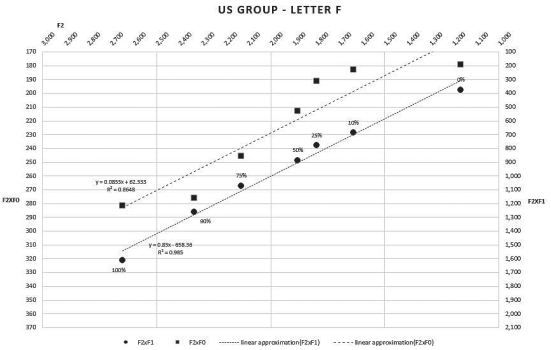

Fig 36. Letter 'F' – Group US

3.7 Letter ‘G’

The triplet of the fitting linear curves were measured as (1.0891, -1501.1, 0.9022) for the group JP (see Fig.37), (1.0397, -1390.8, 0.8523) for S1 (Fig. 38), (1.1451, -1614.9, 0.9508) for E1 (Fig. 39), (1.0676, -1463.9, 0.8938) for E2 (Fig. 40), (1.0447, -1707.9, 0.7919) for UK (Fig. 41), and (0.8334, -1140.9, 0.6585) for US as depicted in Fig. 42. It is clear from this bevy of data that the Japanese students showed relatively strong correlation with linear curves by showing an r-squared greater than 0.85 whereas the native speakers were better characterized by an exponential for the UK group and a polynomial curve for US group. For the groups of students, the slopes had, qualitatively speaking, values akin to each other.

On the other hand, the percentiles of the pitch frequencies read (0.0647, 83.076, 0.9346) for the group JP, and (0.0679, 71.438, 0.926) for S1, (0.0485, 125.38, 0.7719)

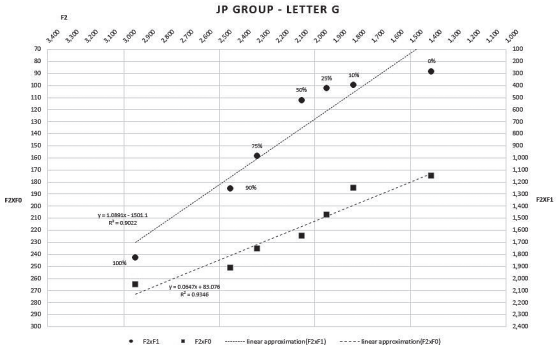

Fig 37. Letter ‘G’ – Group JP

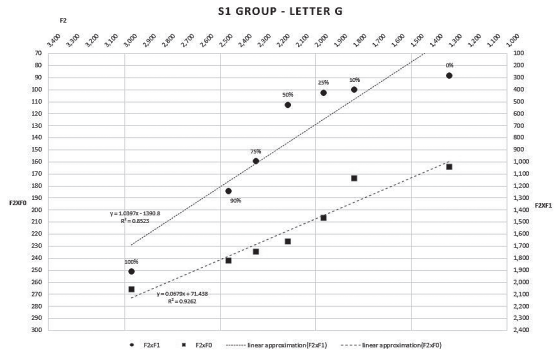

Fig 38. Letter ‘G’ – Group S1

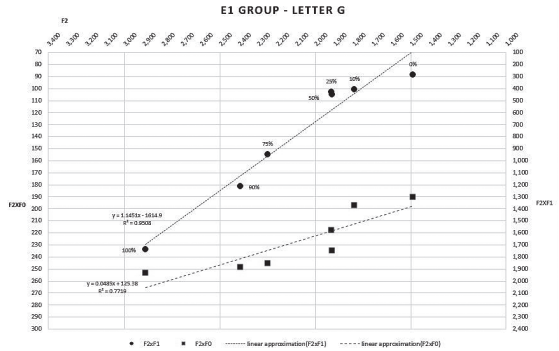

Fig 39. Letter ‘G’ – Group E1

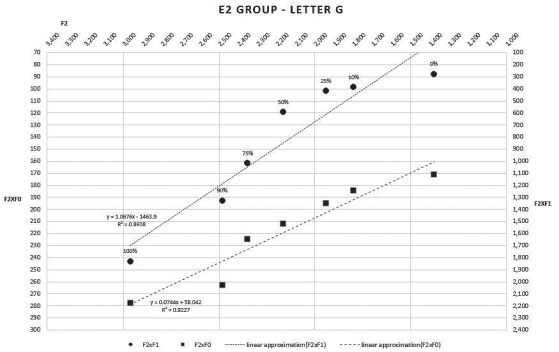

Fig 40. Letter ‘G’ – Group E2

for E1, (0.0744, 58.042, 0.9227) for E2 (0.0369, 114.56, 0.8862) for UK, and (0.0503, 83.645, 0.8975) for US. Thus, E1 had a nonlinear modulation with r-squared value less than 0.75, whereas the other groups were linear correlated with r-squared values greater than 0.85. Still, the slope values suggest that the students in general had a pitch modulation closer to the group US than UK.

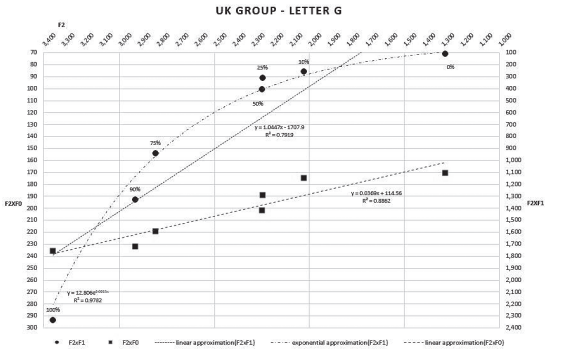

Fig 41. Letter 'G' – Group UK

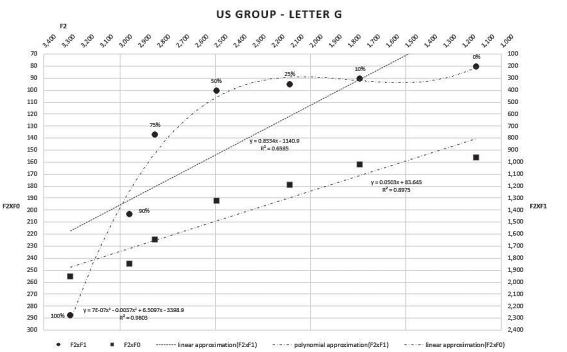

Fig 42. Letter 'G' – Group US

3.8 Letter 'H'

Figure 43 shows that the group of Japanese students had a linear curve having fitting characteristics yielded by the triplet (1.070, -1288.3, 0.9498) for the group JP, whereas Fig. 44 through 46 show that the group S1 had (1.0589, -1285.5, 0.9315), E1 scored (1.0554, -1196.7, 0.9566) and E2 measured (1.0835, -1363.7, 0.9519). As for the group of native speakers, the triplet read (1.3034, -2046.1, 0.9365) for the group UK as seen in Fig. 47 and (0.9618, -1176.7, 0.8714)

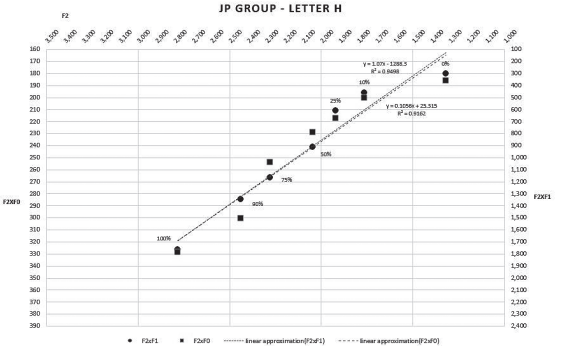

Fig 43. Letter 'H' – Group JP

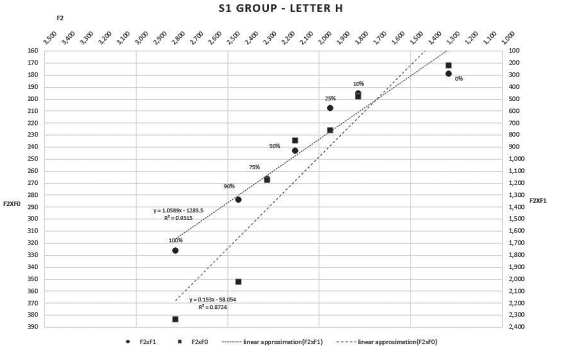

Fig 44. Letter 'H' – Group S1

for the group US (Fig. 48). Since the r-squared values of all these curves were greater than 0.85, the linear correlations can be taken to be relatively high, so that the tongue positioning and the rounding of the lips can be seen as proportional one to another. Yet, the slopes of the lines suggest that the formant variations of the students were between UK and US.

Analyses of the pitch percentiles show that the group JP had (0.1056, 25.315, 0.9162); S1, (0.153, -58.054, 0.8724); E1, (0.0989, 54.406, 0.9357); E2, (0.0572, 92.504, 0.8993); UK, (0.0409, 125.66, 0.911); US, (0.0505, 97.21, 0.9703). The main features figured out from these results are the strong linearity of the fitting curves and the relatively higher slope values for the students compared to the native speakers. In other words, the pitch frequencies increased smoothly along the time axis and the ratio of variations were large for the students.

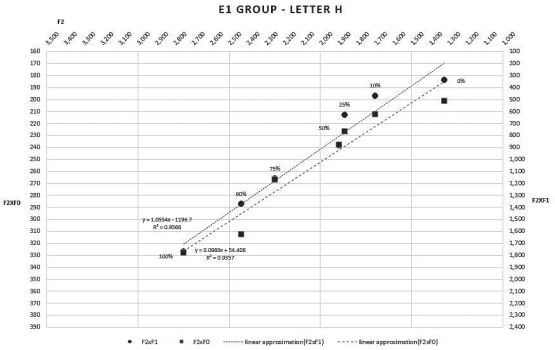

Fig 45. Letter ‘H’ – Group E1

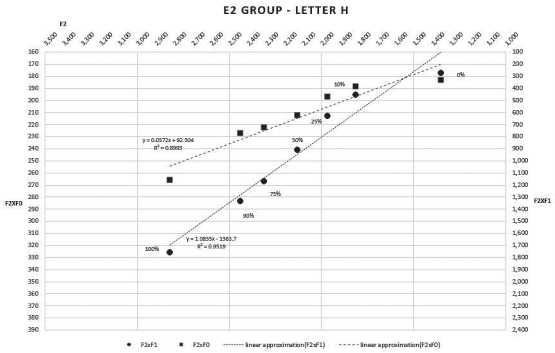

Fig 46. Letter ‘H’ – Group E2

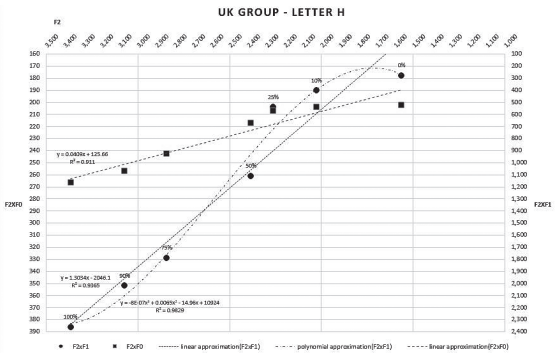

Fig 47. Letter ‘H’ – Group UK

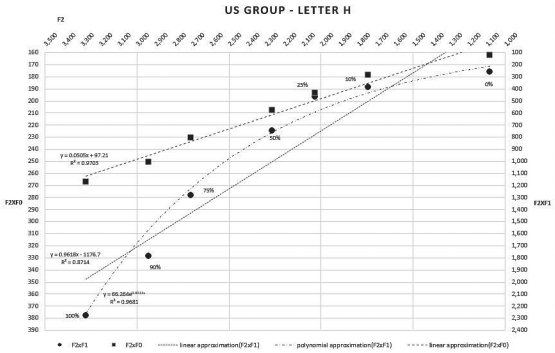

Fig 48. Letter ‘H’ – Group US

3.9 Letter 'I'

Figure 49 shows that the group JP had a fitting curve characterized by the triplet (0.7909, -604.33, 0.9678), whereas Fig. 50 gives (0.8167, -684.09, 0.9545) for S1, Fig. 51 provides (0.7822, -567.29, 0.9554) for E1, Fig. 52 displays (0.7504, -516.22, 0.9568) for E2, Fig. 53 depicts (0.5661, -170.9, 0.8703) for UK, and Fig. 54 says that the values were (0.5865, -254.85, 0.9489) for US. The r-squared values of the lines were all greater than 0.85, indicating a highly linear correlation, which means proportional and synchronized movements of the tongues and lips. Furthermore, the values of the slopes were higher for the students when compared to the native speakers. In fact, contrary to the slope values lesser than 0.6 of the natives, the students had values greater than 0.75.

The analyses of the pitch percentiles gave (0.0455, 131.72, 0.8967) for the group JP, (0.0649, 97.293, 0.9002) for S1, (0.032, 165.64, 0.8048) for E1, (0.0411,

Fig 49. Letter 'I' – Group JP

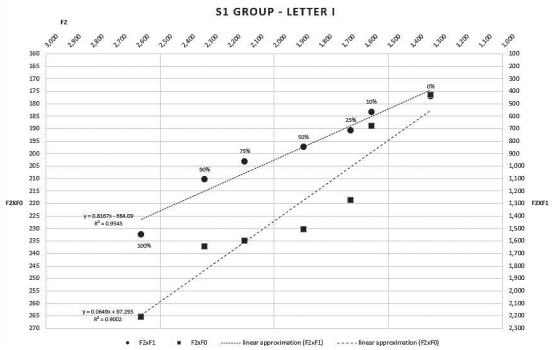

Fig 50. Letter 'I' – Group S1

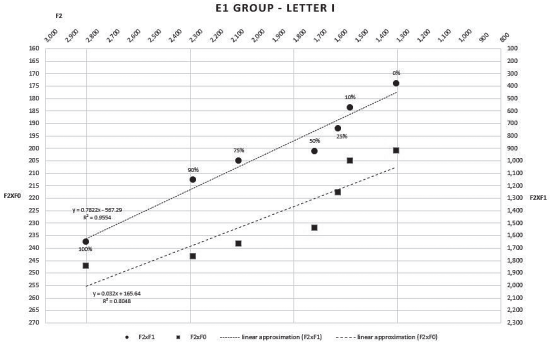

Fig 51. Letter 'I' – Group E1

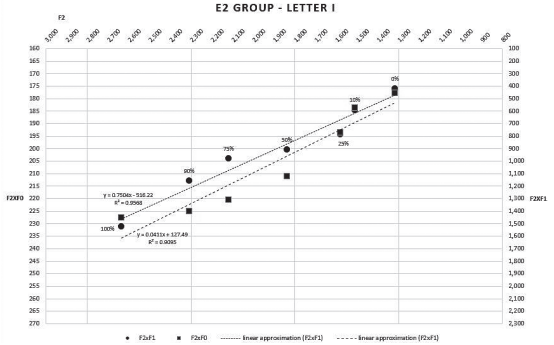

Fig 52. Letter 'I' – Group E2

127.49, 0.9095) for E2, (0.032, 133.22, 0.9225) for UK, and (0.0437, 111.56, 0.985) for US. The values of the r-squared terms say that linear curves fitted pretty well on the percentiles. In addition, leaving out the group S1, the graphs show that the slopes of the students had values more or less around the values measured for the native speakers. Thus, the pitch frequencies varied temporally in a linear fashion with values increasing similarly for all groups.

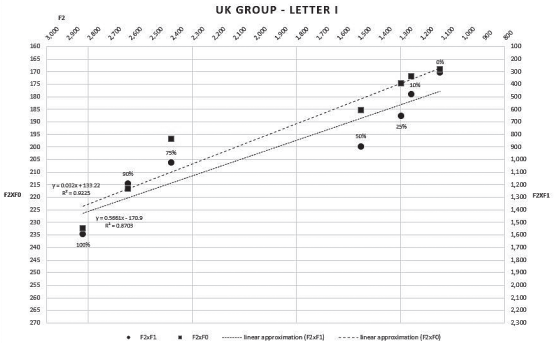

Fig 53. Letter ‘I’ – Group UK

Fig 54. Letter ‘I’ – Group US

3.10 Letter ‘J’

The processing results for this letter are represented in figures 55 through 60. They show that the formant percentiles of the group JP can be characterized by the fitting curve given by the triplet (1.0621, -1357.7, 0.937), the group S1 by (0.9592, -1153.9, 0.918), the group E1 by (1.15, -1516.8, 0.96), the group E2 by (1.069, -1378.0, 0.9000), the group UK by (1.2787, -2037.4, 0.8906), and the group US by (0.9045, -1146, 0.8001).

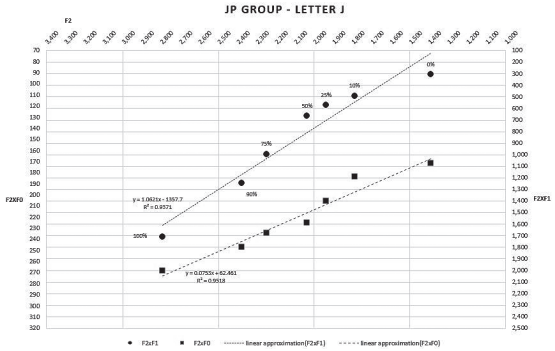

Fig 55. Letter ‘J’ – Group JP

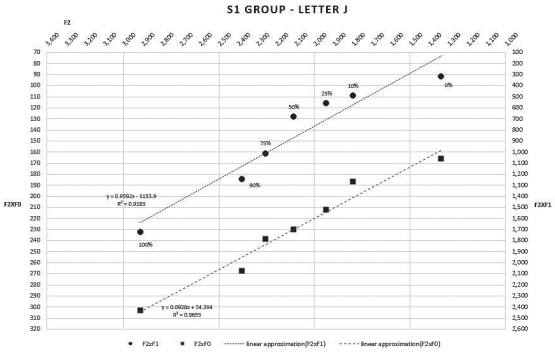

Fig 56. Letter ‘J’ – Group S1

Nevertheless the r-squared of the native speakers had satisfactory values meaning high correlation with linear curves, the fitting curves would be better described by a nonlinear curve; namely an exponential and a second order polynomial, as shown in figures 59 and 60. As for the values of the slopes, the students had values in-between the groups UK and US.

On the other hand, the pitch percentiles translated into (0.0753, 62.461, 0.9518) for the group JP, (0.0928, 34.294, 0.9653) for S1, (0.0637, 95.178, 0.8941) for E1, (0.0637, 67.132, 0.9011) for E2, (0.0506, 92.443, 0.9171) for UK, and (0.0546, 72.679, 0.9425) for US. The fitting curves were all linear with r-squared values greater than 0.85. Yet, comparing the slopes of the students against the natives, the former had values relatively larger than the latter, meaning that the ratio of variation was greater.

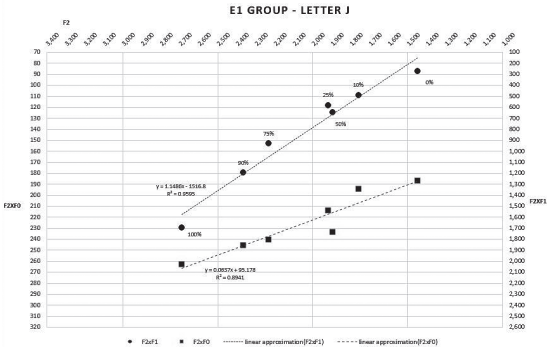

Fig 57. Letter 'J' – Group E1

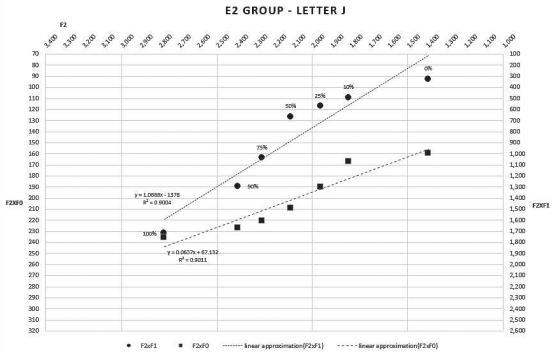

Fig 58. Letter 'J' – Group E2

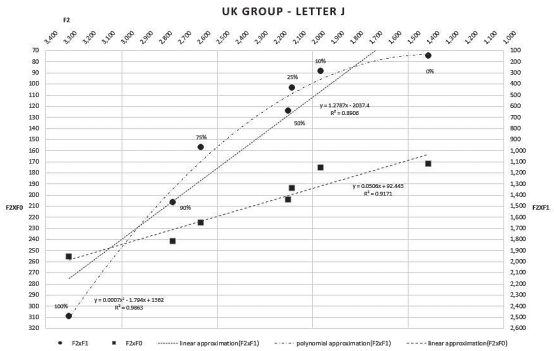

Fig 59. Letter 'J' – Group UK

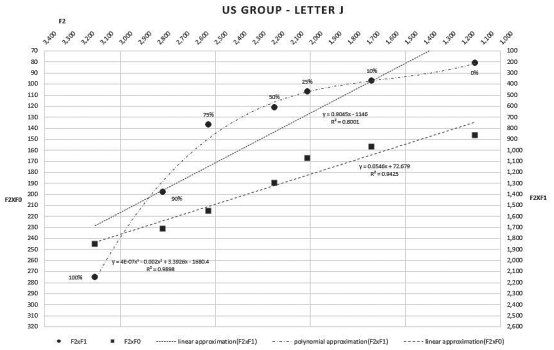

Fig 60. Letter 'J' – Group US

3.11 Letter ‘K’

The triplet characterizing the linear fitting curves of the formants percentiles read (0.9486, -1129.5, 0.8964) for the group JP, (0.9073, -1038, 0.8565) for S1, (0.9229, -1075.4, 0.9043) for E1, (0.993, -1232, 0.8926) for E2, (1.0138, -1472.5, 0.7908) for UK, and (0.6261, -629.9, 0.8149) for US. Looking at the r-squared values, one sees that the students had values greater than 0.89 whereas the natives yielded values smaller than 0.82. This fact suggests that, in contrast to the formants of the natives, which were modulated following respectively a polynomial and an exponential curve, the students varied the formant frequencies constantly over the time line.

The analyses of the pitch percentiles revealed that the triplet for the group JP was (0.0841, 53.668, 0.9428), for the group S1 it read (0.0821, 59.769, 0.9638), for E1 the set was (0.0835, 64.516, 0.9413), for the group E2 the values were (0.0854, 38.25, 0.8631), for the group UK the trio translated into

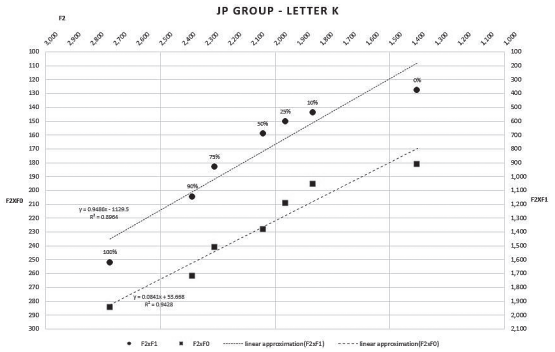

Fig 61. Letter ‘K’ – Group JP

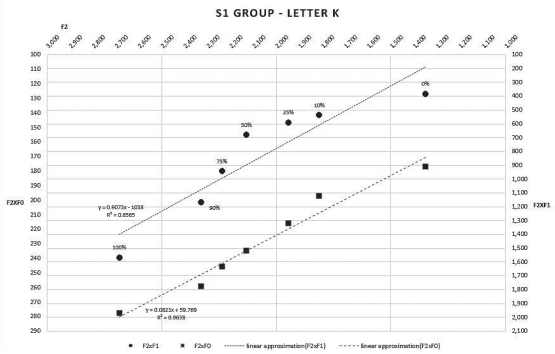

Fig 62. Letter ‘K’ – Group S1

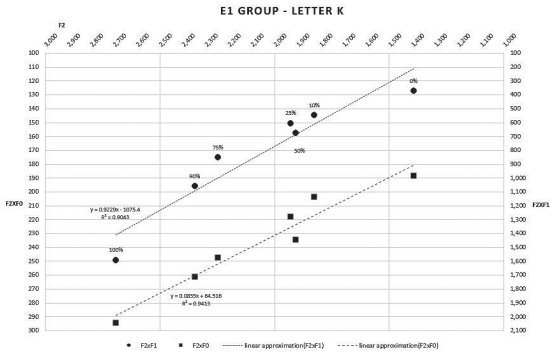

Fig 63. Letter ‘K’ – Group E1

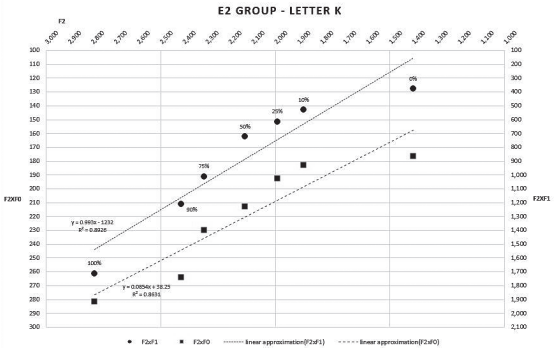

Fig 64. Letter ‘K’ – Group E2

(0.0718, 56.204, 0.7458), and for the group US the numbers were (0.0657, 56.433, 0.9051). From the r-squared values, it is clear that UK would be better expressed by a nonlinear curve rather than a line. As for the slopes, the students had values greater than those measured for the group US, which means larger temporal variations during the production of the utterances.

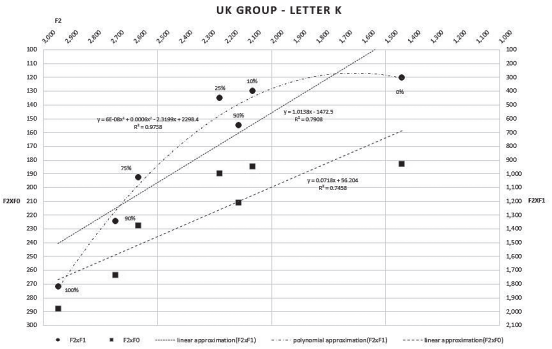

Fig 65. Letter 'K' – Group UK

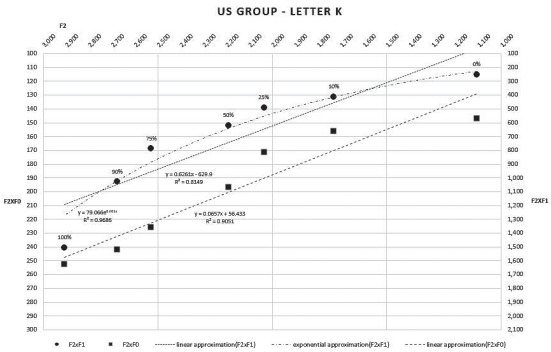

Fig 66. Letter 'K' – Group US

3.12 Letter 'L'

Figures 67 through 72 depicts the approximation curves for the percentiles of the formants and the pitch frequencies. For the formants, the (slope, intercept, r-squared) values were computed as (0.8647, -835.23, 0.9697) for the group JP, (0.9335, -1014.5, 0.929) for the group S1, (0.7781, -651.64, 0.9696) for the group E1, (0.8717, -821.99, 0.9746) for E2, (0.6335, -242.26, 0.9553) for the group UK, and (0.5265, -88.829, 0.9803) for the group US. The values of r-squared were all greater than 0.9 meaning that the formant percentiles

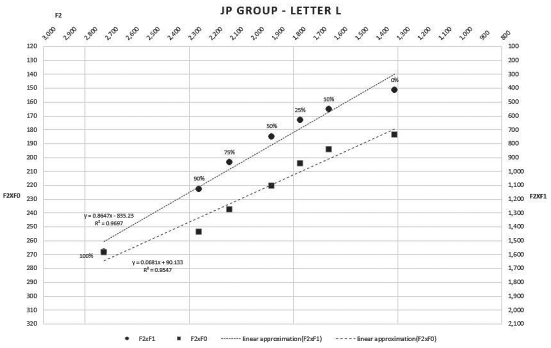

Fig 67. Letter 'L' – Group JP

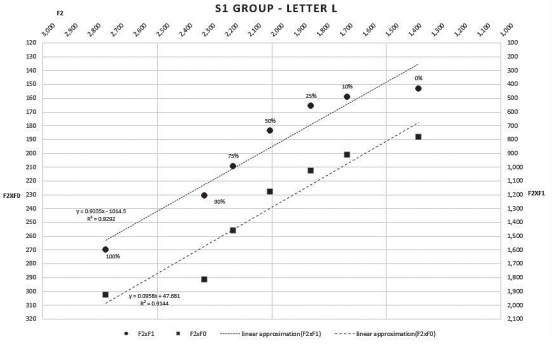

Fig 68. Letter 'L' – Group S1

were fairly well represented by straight lines. As far as the slopes are concerned, the values of the students were greater than 0.75. Actually, this number is a bit higher than the slopes of the native speakers, which had the largest value for the UK group, marking 0.63.

Now, spotting the pitch percentiles, the figures show that the fitting curves were described by (0.0681, 90.133, 0.9547) for JP, (0.0958, 47.681, 0.9144) for S1, (0.0408, 147.59, 0.8903) for E1, (0.0666, 72.238, 0.9761) for E2, (0.0358, 152.25, 0.9655) for UK, (0.0541, 110.7, 0.9203) for US. Since all the r-squared values of the groups were greater than 0.9, the linear variation can be taken as one of the main characteristics of the results. Another point is that the students had steeper variations than the natives as seen from the slope values.

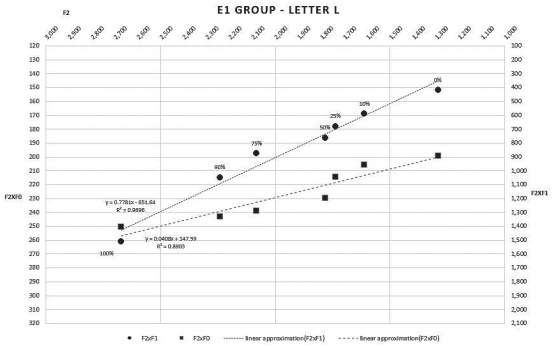

Fig 69. Letter 'L' – Group E1

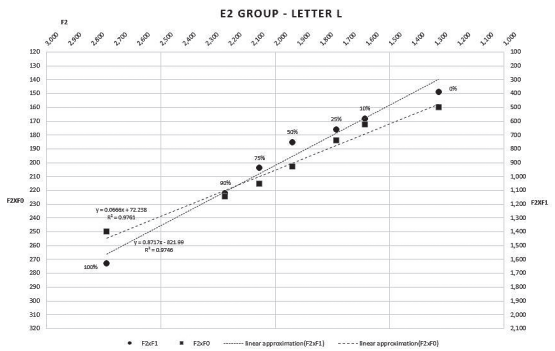

Fig 70. Letter 'L' – Group E2

Fig 71. Letter 'L' – Group UK

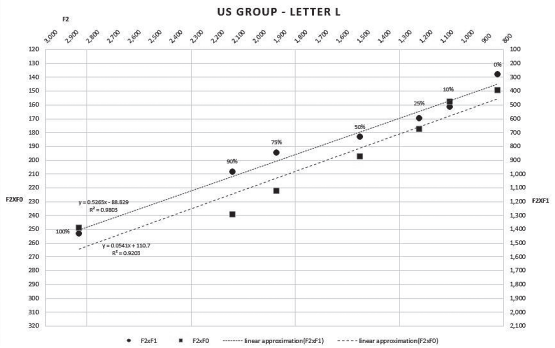

Fig 72. Letter 'L' – Group US

3.13 Letter 'M'

As shown in Fig 73, the group JP had a fitting curve with the triplet defined by (1.0033, -1153, 0.9856), whereas S1 by (0.9379, -1058.7, 0.9659), E1 by (0.9466, -1016.9, 0.9735), E2 by (1.1325, -1403.9, 0.9959), UK by (1.1495, -1451.8, 0.9694), and US by (0.8517, -839.64, 0.9652). The strong linear behaviors of the formants are guaranteed for these approximations by the large numbers computed out for the r-squared values, which were all greater than 0.95. Now, the slopes of the approximants were such that the values of the students were upper and lower bounded by the values of the slopes calculated for the native speakers.

Now, as for the percentiles of the pitch frequencies, the curve fitting computations led to (0.0848, 57.03, 0.959) for the group JP, (0.084, 62.94, 0.9311) for the group S1, (0.052, 123.72, 0.9669) for the group E1, (0.1268, -35.992, 0.8625) for E2, (0.0413, 124.03, 0.9126) for the group UK, and

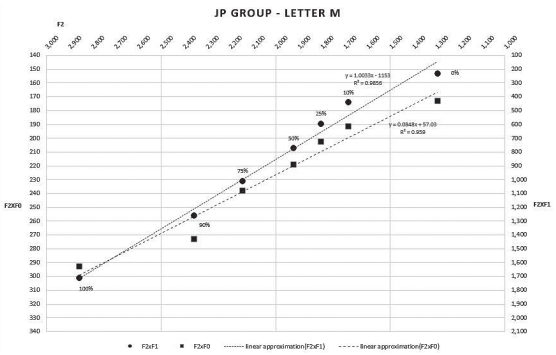

Fig 73. Letter 'M' – Group JP

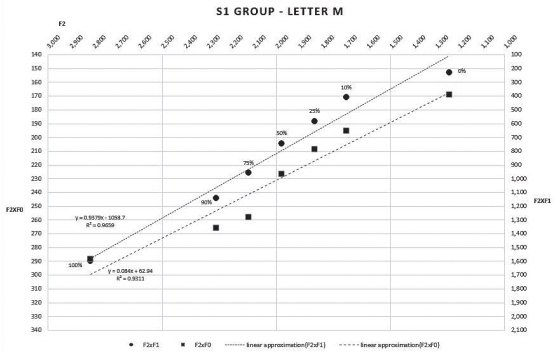

Fig 74. Letter 'M' – Group S1

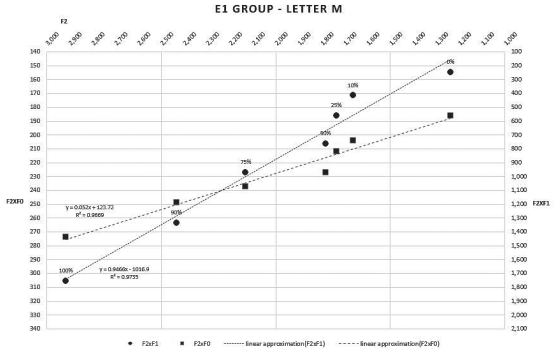

Fig 75. Letter 'M' – Group E1

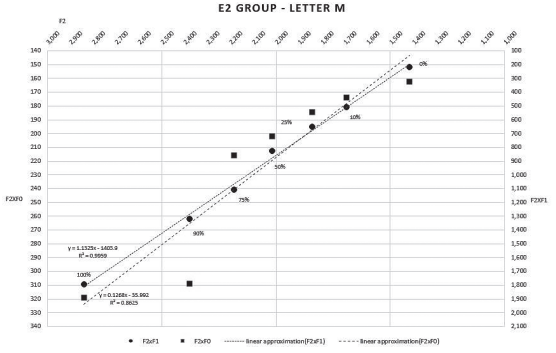

Fig 76. Letter 'M' – Group E2

(0.0658, 80.645, 0.9028) for US. It is straightforward that, excluding the group E2 which had r-squared value scored 0.86, the remaining groups had all r-squared values greater than 0.9, endorsing the linear model to represent the temporal variation of the percentiles. Another point to note is that the groups of students had different slope values so that some were greater than those of the natives, and others were smaller.

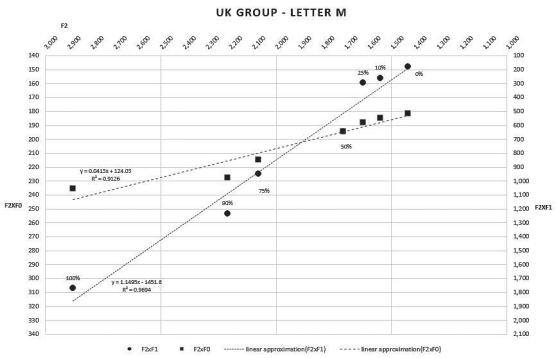

Fig 77. Letter ‘M’ – Group UK

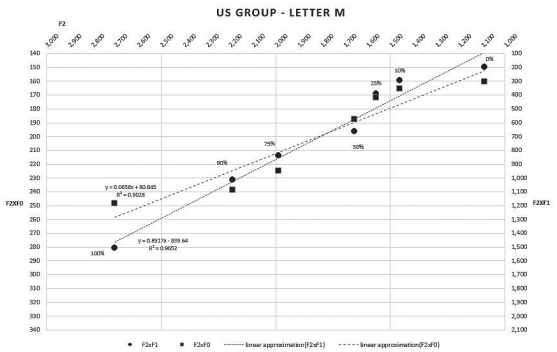

Fig 78. Letter ‘M’ – Group US

3.14 Letter ‘N’

From Fig 79, the fitting curve of the formant percentiles for the group of students as a whole, namely JP, is given by the triplet (0.9993, -1163.3, 0.9704), whereas its subgroups were typified by the triplet (0.9332, -1057.9, 0.9567) for the group S1 (Fig 80), (0.9542, -1028.2, 0.9265) for the group E1 (Fig 81), (1.0995, -1394.6, 0.9811) for the group E2 (Fig 82). Now focusing on the groups of native speakers, the group UK had (0.8762, -907.73, 0.9063),

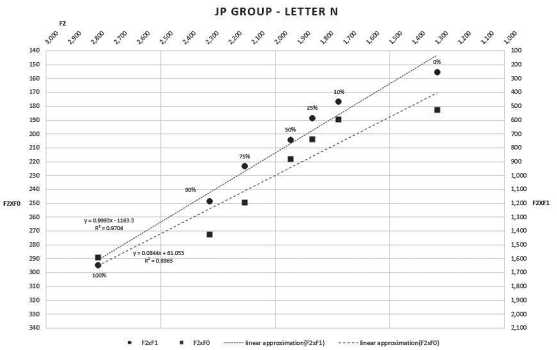

Fig 79. Letter ‘N’ – Group JP

Fig 80. Letter ‘N’ – Group S1

whereas US was represented by the fitting curve established by (0.7063, -594.62, 0.8941). As all the r-squared values were greater then 0.85, it is reasonable to assume the linearity of sound production in terms of mouth movement coordination. Note also that the slopes of the students had values relatively greater than those computed for the native speakers.

For the percentiles of the pitch frequencies, the group JP had fitting curves defined by the triplet (0.0844, 61.053, 0.8963), S1 by (0.0757, 76.427, 0.8801), E1 by (0.0507, 129.67, 0.9147), E2 by (0.1349, -45.542, 0.8338), UK by (0.0549, 95.023, 0.8241), and US by (0.0587, 95.188, 0.9126). Thus, JP, S1, E1, and US had r-squared values greater than 0.85 whereas E2 and UK scored lower. In addition, unlike other groups, the slopes of E2, UK and US were in-between 0.05 and 0.06.

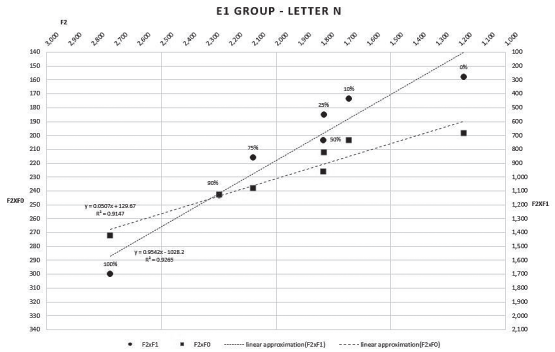

Fig 81. Letter 'N' – Group E1

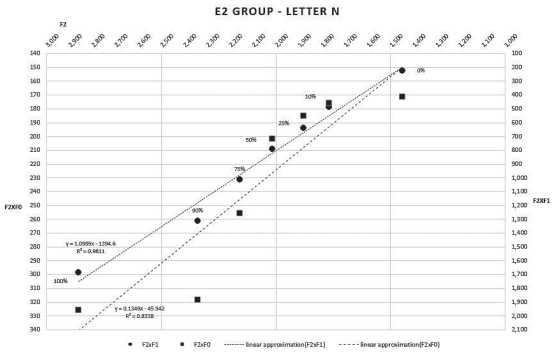

Fig 82. Letter 'N' – Group E2

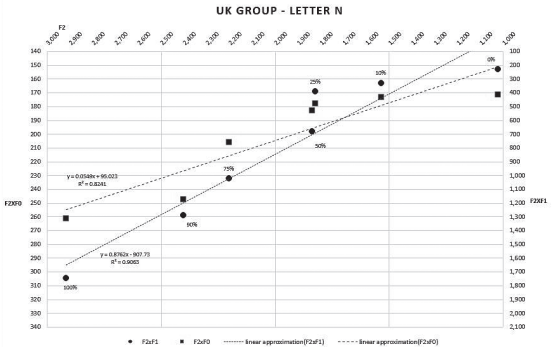

Fig 83. Letter 'N' – Group UK

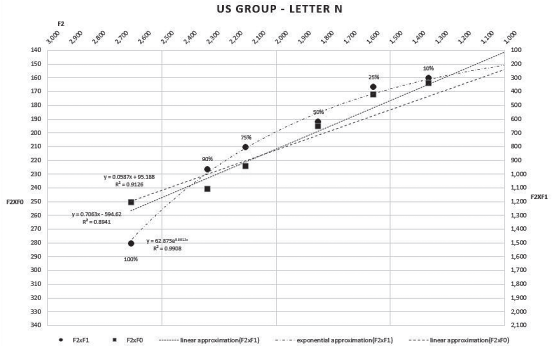

Fig 84. Letter 'N' – Group US

3.15 Letter ‘O’

The results for this letter are displayed in figures 85 through 90. First, analyzing the formant percentiles, Fig 85 shows that the fitting curve for the group JP was given by (0.5322, -22.812, 0.9489), and the other figures say that S1 was approximated by (0.4803, 66.698, 0.9747), E1 by (0.5712, -80.208, 0.9177), E2 by (0.541, - 49.061, 0.9471), UK by (0.8997, -884.46, 0.996), and US by (0.5228, -49.04, 0.9959). Taking the natives as reference - whose formants were linearly varying parameters - it is obvious that the students also behaved similarly. Comparing the slopes of the students and the natives, we see that the students had values closer to the group US, which had a value around 0.5 whereas the UK group orbited around 0.8.

Next, analyzing the percentiles of the pitch frequencies, the results showed that the group JP was expressed by the triplet gathered as (0.046, 138.9, 0.88), S1 as (0.0689, 107, 0.8484), E1 as (0.0362, 163.1, 0.8174), E2

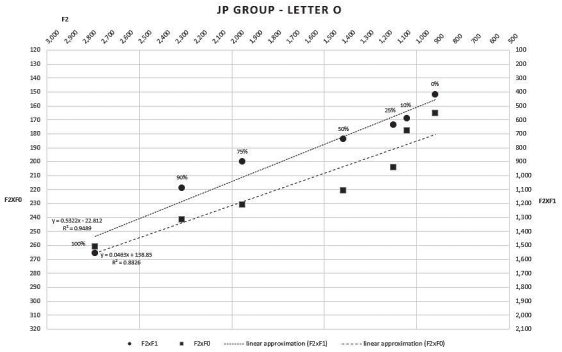

Fig 85. Letter ‘O’ – Group JP

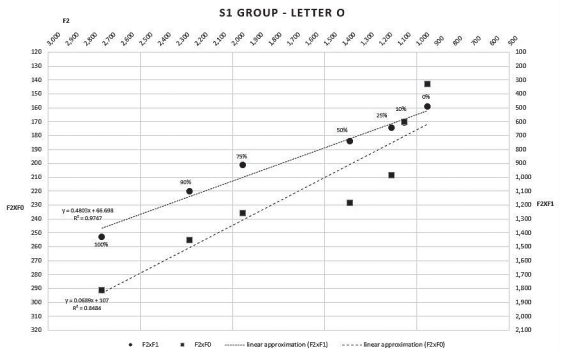

Fig 86. Letter ‘O’ – Group S1

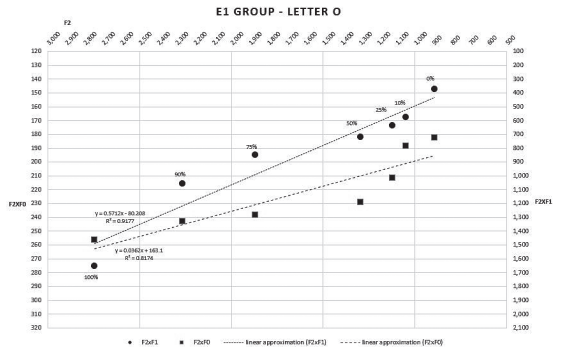

Fig 87. Letter ‘O’ – Group E1

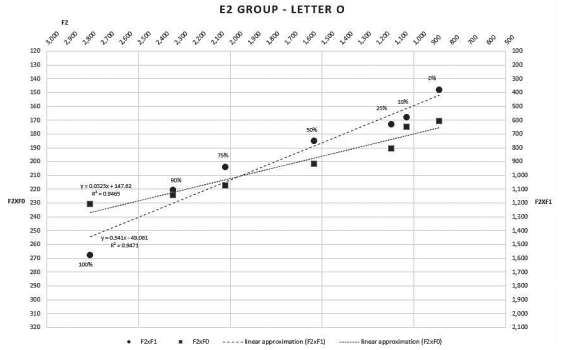

Fig 88. Letter ‘O’ – Group E2

as (0.0323, 147.62, 0.9465), UK as (0.0494, 111.74, 0.8908), and US as (0.0508, 124.08, 0.9016). Hence, S1 and E1 had r-squared values between 0.80 and 0.85, whereas the other groups were larger than 0.85. Moreover, the slopes varied across the groups of students, so that a comparison with the natives' numbers would not give us any striking characteristics.

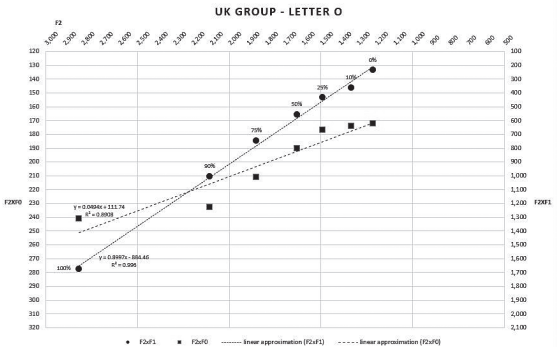

Fig 89. Letter 'O' – Group UK

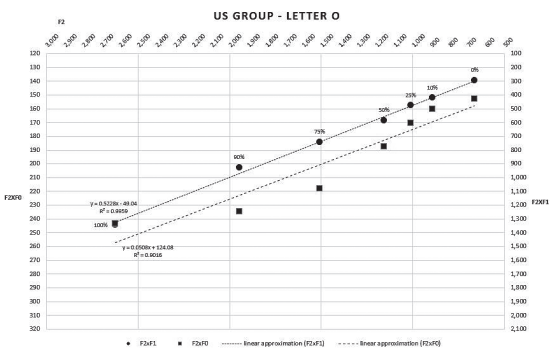

Fig 90. Letter 'O' – Group US

3.16 Letter 'P'

In terms of the sound classification, this letter is classified as belonging to the same category as the letter 'B', so it was expected that the behavioral patterns of the formant percentiles had some similarities with each other. In fact, the triplet defining the fitting curve for the group JP was computed as (1.1331, -1614.7, 0.9001) as seen in figure 91. As for the group S1, the triplet read (1.1564, -1641.9, 0.9093), for the group E1 the set was (1.1282, -1616.6, 0.9353); E2 gave (1.078, -1505.4, 0.8272),

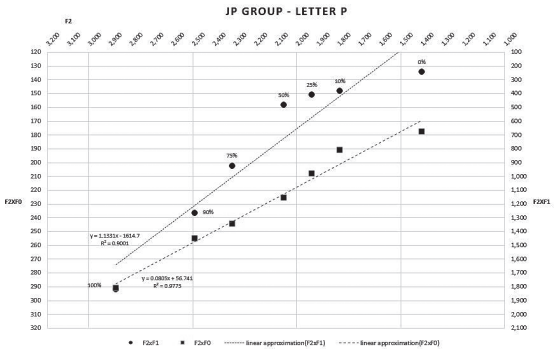

Fig 91. Letter 'P' – Group JP

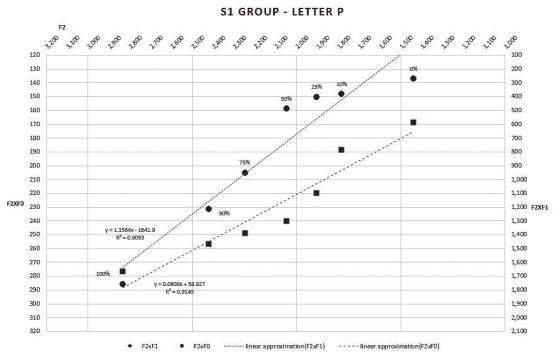

Fig 92. Letter 'P' – Group S1

UK yielded (1.2162, -2135.5, 0.6713), and US (0.6748, -826.45, 0.7132). The r-squared values of the natives suggest that the formants were better modelled by nonlinear curves, which means that the mouth made complex movements during the utterance generation. In contrast, the students moved the tongues and the lips in a proportional way.

For the percentiles of the pitch frequencies, the triplet of the group JP was (0.0805, 56.741, 0.9773); S1, (0.0809, 58.927, 0.9149); E1, (0.0748, 71.528, 0.9626); E2, (0.0799, 51.219, 0.8339); UK, (0.0924, 4.3142, 0.7957); US, (0.0953, 4.7605, 0.893). The group of students and US had r-squared values greater than 0.85, supporting the linear model, whereas UK went lower than 0.80. Yet, the slopes of the students were around 0.08, which were lesser than the value measured for US, which scored 0.0953.

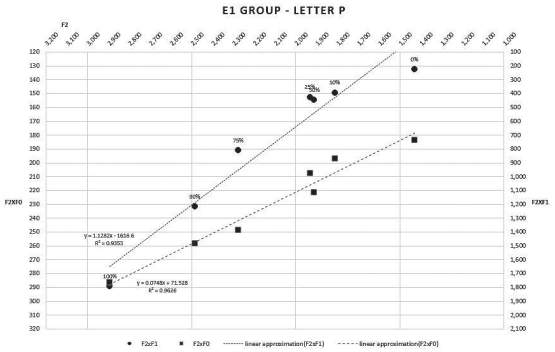

Fig 93. Letter ‘P’ – Group E1

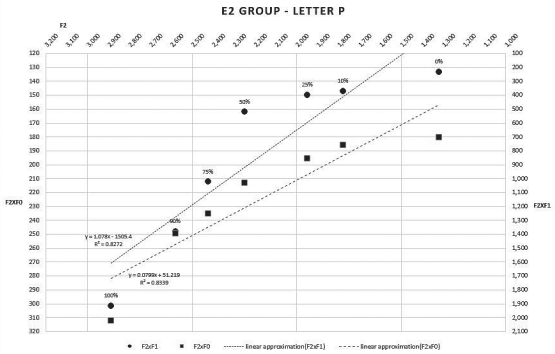

Fig 94. Letter ‘P’ – Group E2

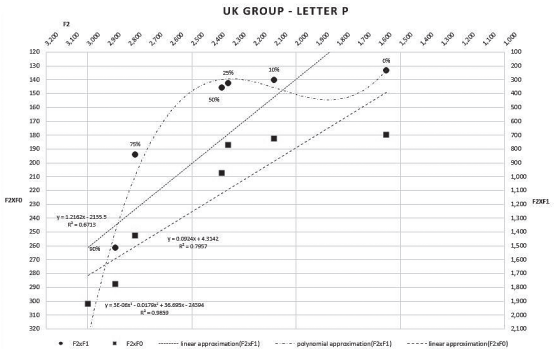

Fig 95. Letter ‘P’ – Group UK

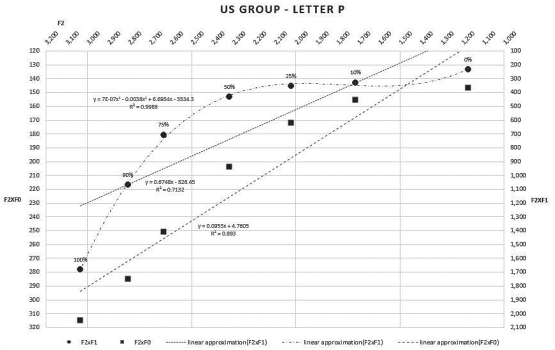

Fig 96. Letter ‘P’ – Group US

3.17 Letter 'Q'

The following triplets for the formant percentiles were obtained here: (1.1185, -1449.3, 0.9757) for the group JP, (1.1215, -1436.2, 0.9713) for the group S1, (1.1263, -1513.4, 0.9809) for the group E1, (1.0962, -1372.3, 0.9679) for the group E2, (1.0816, -1299.6, 0.9467) for the group UK, and (0.6261, -489.32, 0.8868) for the group US. Since the values of the r-squared were all greater than 0.85, the linear model expressed reasonably well the movements of the tongue and lips, which were proportional to each other. As for the slopes, ruling out the group US, which scored 0.6261, the other groups were all around 1.1. Thus, in terms of the production of the formants, the students had higher affinity to the group UK.

Now, for the percentiles of pitch frequencies, the computations output the following triplets: (0.0821, 59.715, 0.9188) for the group JP, (0.0848, 57.306, 0.8641), for S1, (0.0716, 85.229, 0.8209) for

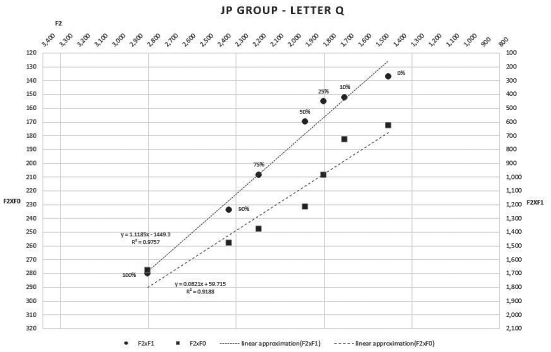

Fig 97. Letter 'Q' – Group JP

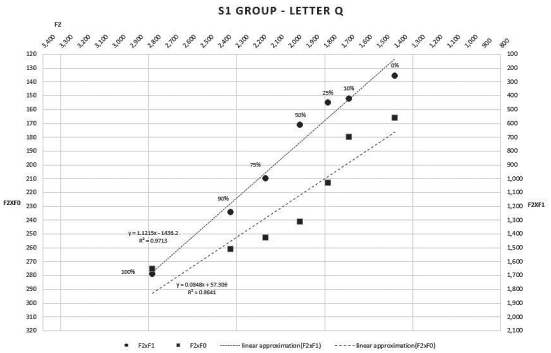

Fig 98. Letter 'Q' – Group S1

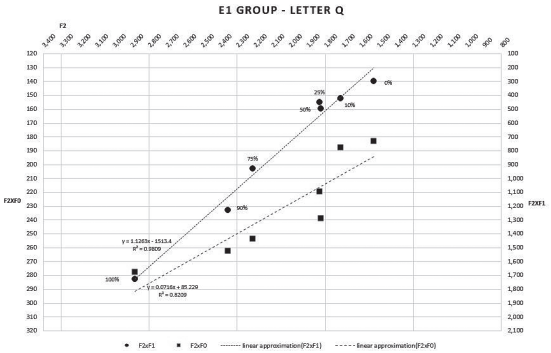

Fig 99. Letter 'Q' – Group E1

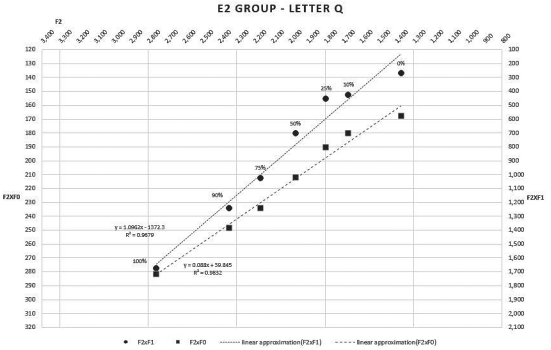

Fig 100. Letter 'Q' – Group E2

E1, (0.088, 39.845, 0.9832) for E2, (0.0365, 154.53, 0.9683) for UK, and (0.068, 89.383, 0.9626) for US. From the r-squared values, it is fair to assume that the linear model explains very well the temporal variation of the percentiles. Yet, the slopes show that the group of students, which were greater than 0.07, had steeper variations compared to the natives (0.0365 for UK, and 0.068 for US.)

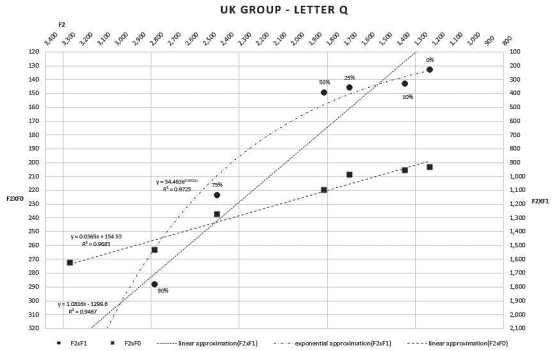

Fig 101. Letter ‘Q’ – Group UK

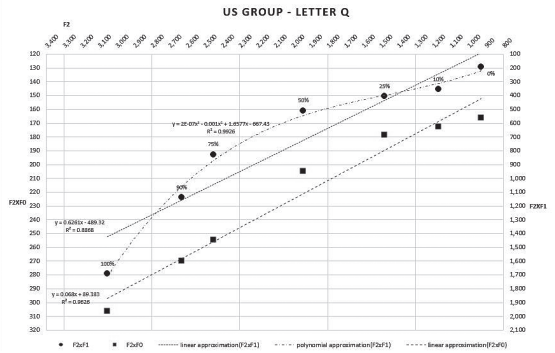

Fig 102. Letter ‘Q’ – Group US

3.18 Letter ‘R’

This letter is amongst the sounds whose production is very different from the sounds embedded in the Japanese language. The results are shown in figures 103 through 108. For the percentiles of the formants, the calculations rendered the triplets (0.7712, -464.25, 0.9291) for the group JP, (0.9283, -799.86, 0.9638) for the group S1, (0.7334, -381.81, 0.8566), for the group E1, (0.6364, -188.9, 0.9285) for the group E2, (0.5997, -50.284, 0.9392) for the group UK, and (0.6921,

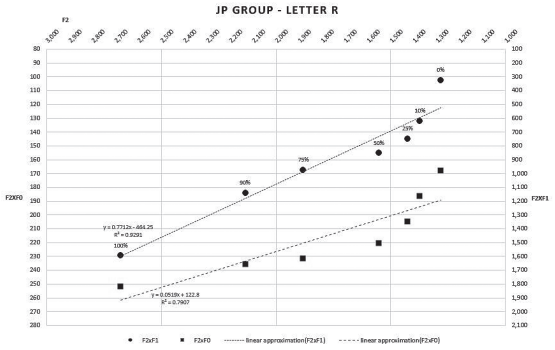

Fig 103. Letter ‘R’ – Group JP

Fig 104. Letter ‘R’ – Group S1

-282.82, 0.9791) for the group US. The analyses of the fitness of the linear model on the basis of r-squared values show that all the groups are very well represented by the linear curves described so far. In addition, the analyses of the variation rates suggest that the students in general had larger values than the native speakers. Note that the group E2 had an r-squared value between the groups UK and US.

As far as the percentiles of the pitch frequencies are concerned, the triplet read (0.0519, 122.8, 0.7907) for the group JP, (0.0617, 102.69, 0.7042) for S1, (0.0583, 123.71, 0.8212) for E1, (0.0361, 139.39, 0.8178) for E2, (0.0408, 140.98, 0.935) for UK, and (0.0666, 86.144, 0.8711) for the group US. The linear model fitted well for the native speakers; however, the r-squared values for the students indicate that the level of linearity was not as strong as seen in the natives.

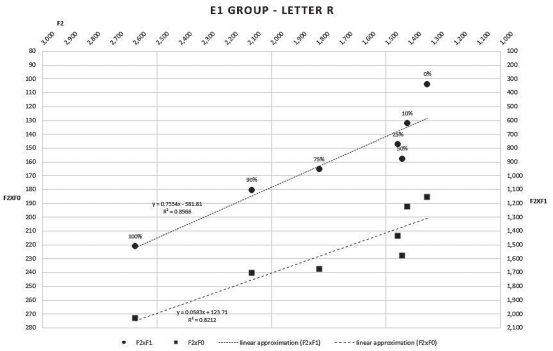

Fig 105. Letter ‘R’ – Group E1

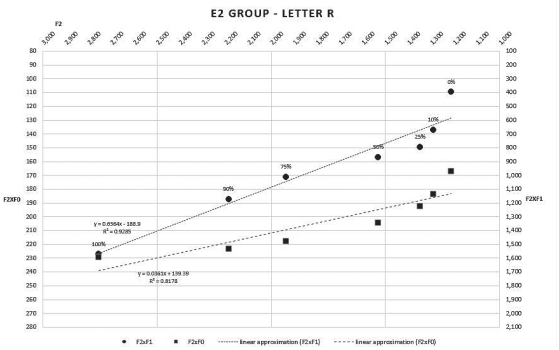

Fig 106. Letter ‘R’ – Group E2

Fig 107. Letter ‘R’ – Group UK

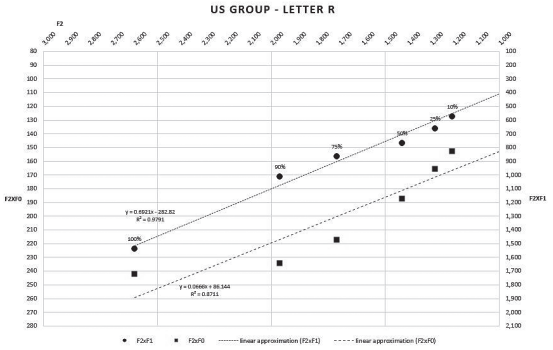

Fig 108. Letter ‘R’ – Group US

3.19 Letter ‘S’

Fig 109 shows that the group JP had a fitting curve for the percentiles of the formants defined by the triplet (1.0757, -1197.9, 0.9505) whereas the group S1 (Fig. 110) as expressed by (1.1121, -1312.4, 0.9245), the group E1 (Fig. 111) described by (0.9348, -850.98, 0.9398), the group E2 (Fig. 112) by (1.2132, -1503.1, 0.9772), the group UK (Fig. 113) by (1.0808, -1245, 0.9576), and the group US (Fig. 114) by (0.7531, -556.24, 0.9858). The values of the r-squared terms unveil that all the groups are well described by the linear models. Leaving out E2, the groups of the students had all slope values within the range bounded by UK and US.

Focusing on the percentiles of the pitch frequencies, the calculation of the approximants led to the triplet (0.0893, 73.862, 0.8066) for the group JP, (0.0882, 78.014, 0.7136) for the group S1, (0.1048, 71.411, 0.7694) for the group E1, (0.0653, 88.488, 0.7701) for the group E2, (0.0902, 29.173, 0.9076) for the

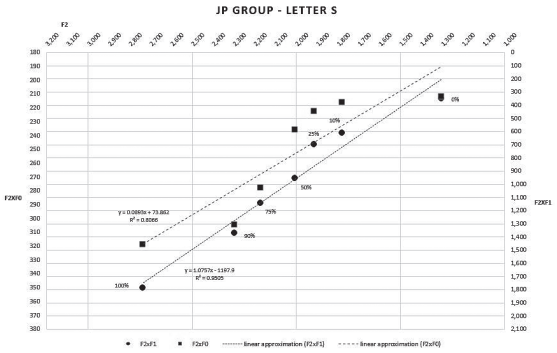

Fig 109. Letter ‘S’ – Group JP

Fig 110. Letter ‘S’ – Group S1

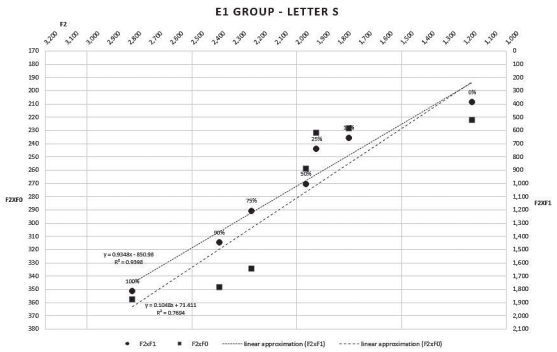

Fig 111. Letter ‘S’ – Group E1

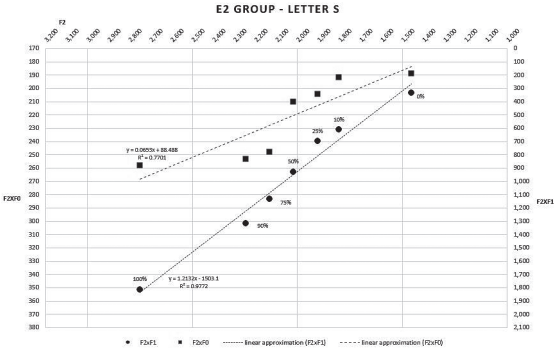

Fig 112. Letter ‘S’ – Group E2

group UK, and (0.0753, 76.708, 0.8781) for the group US. The values of r-squared terms assure that the linear models represent pretty well the pitch percentiles of the groups. As for the slopes, JP and S1 had values in-between UK and US, whereas E1, opposite to E2, had a slope greater than the values computed for the native speakers. These suggest that the groups of students had each different kinds of pitch modulations aimed to reach the frequencies of the native sounds.

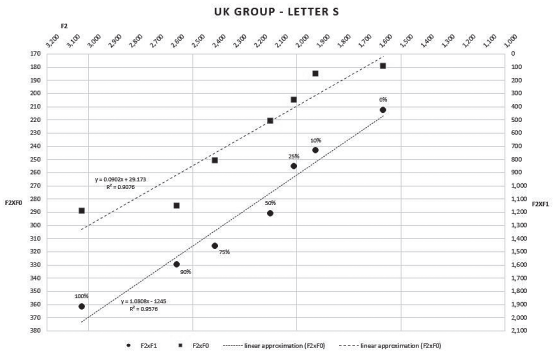

Fig 113. Letter 'S' – Group UK

Fig 114. Letter 'S' – Group US

3.20 Letter 'T'

Figures 115 through 120 show the results of the data processing procedures related to the letter 'T'. For the percentiles of the formants, the approximant for the group JP is determined by the triplet (1.1226, -1583.9, 0.9521), for the group S1 it read (1.1198, -1578.8, 0.9218), for the group E1 the set translated into (1.0872, -1518.9, 0.9477), for the group E2 the numbers were (1.1536, -1639.6, 0.9683),

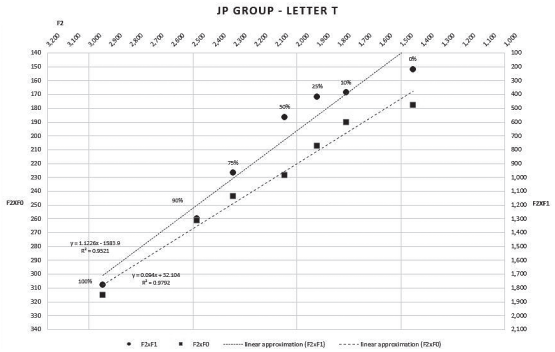

Fig 115. Letter 'T' – Group JP

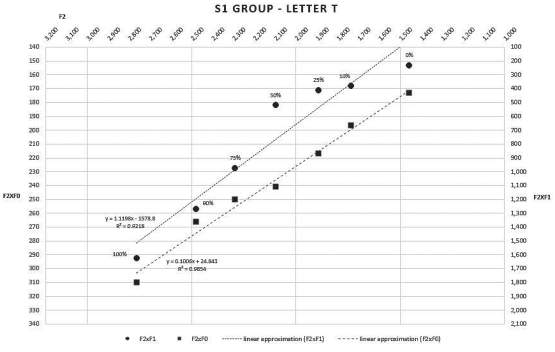

Fig 116. Letter 'T' – Group S1

for the group of natives speakers UK the trio was (0.9606, -1531.1, 0.6584), and finally for the group US the triplet was established as (0.8351, -1169.1, 0.7308). According to the values of r-squared parameters, the linear model fitted well only for the groups of students. The percentiles of the native speakers were better modeled by second order polynomial curves as shown in figures 119 and 120.

On the other hand, the percentiles of the pitch frequencies had approximants given by the triplets (0.094, 32.104, 0.9792) for the group JP, (0.1006, 24.843, 0.9854) for the group S1, (0.0953, 35.343, 0.9513) for the group E1, (0.0858, 35.693, 0.976) for the group E2, (0.0563, 101.95, 0.7203) for the group UK, and (0.0804, 28.444, 0.8826) for the group US. The linear model fitted was endorsed by the r-squared terms.

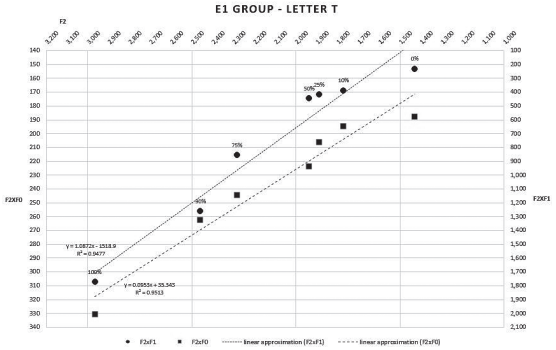

Fig 117. Letter ‘T’ – Group E1

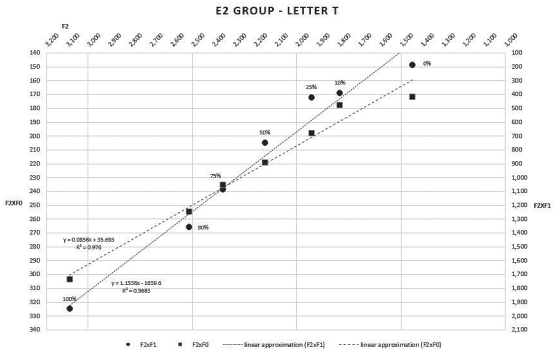

Fig 118. Letter ‘T’ – Group E2

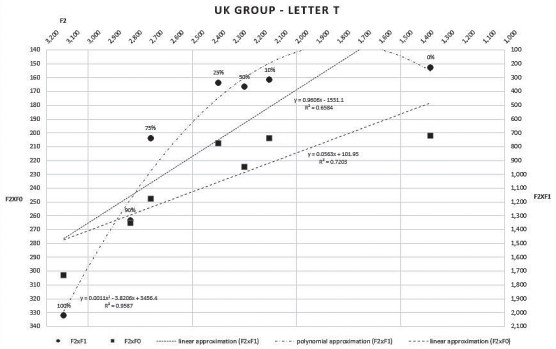

Fig 119. Letter ‘T’ – Group UK

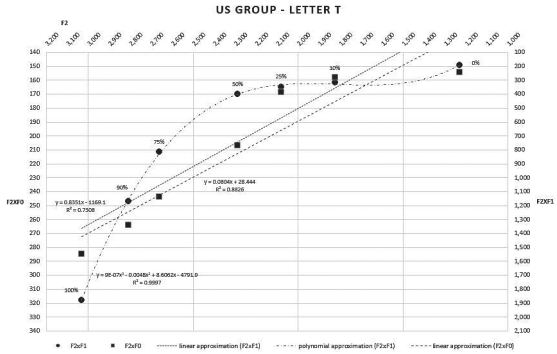

Fig 120. Letter ‘T’ – Group US

3.21 Letter 'U'

The processing of the formant percentiles rendered (1.0622, -1404.1, 0.9663) for the group JP as shown in figure 121, (1.0804, -1433, 0.9668) for the group S1 as depicted in figure 122, (1.0203, -1354.5, 0.938) for the group E1 as illustrated in figure 123, (1.0717, -1392.4, 0.9625) for the group E2 as in figure 124, (0.9031, -1188.1, 0.7063) for the group UK as in figure 125, and (0.527, -386.87, 0.781) for the group US as in figure 126. According to the r-squared values, the groups of students were highly expressed by the linear models, meaning that the movements of the tongue and lips were proportional to each other. On the other hand, the groups of the native speakers had polynomial curves fitting the percentiles better than the linear curves.

The processing of the pitch frequency percentiles output the triplet (0.067, 84.794, 0.926) for the group JP, (0.0928, 37.756, 0.9452) for the group S1, (0.0546, 116, 0.8612) for

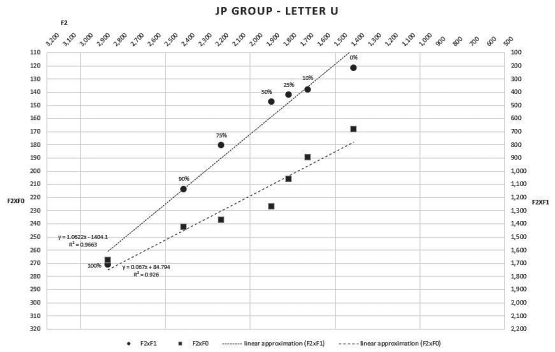

Fig 121. Letter 'U' – Group JP

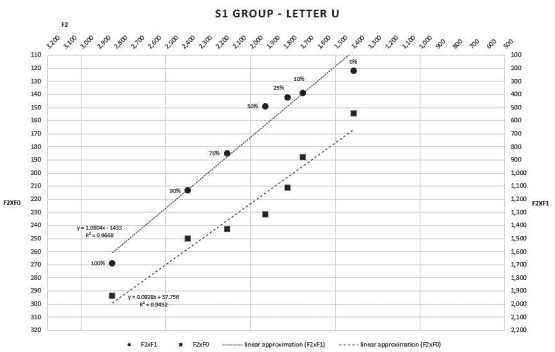

Fig 122. Letter 'U' – Group S1

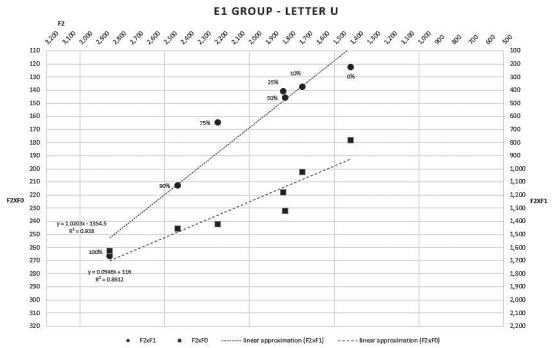

Fig 123. Letter 'U' – Group E1

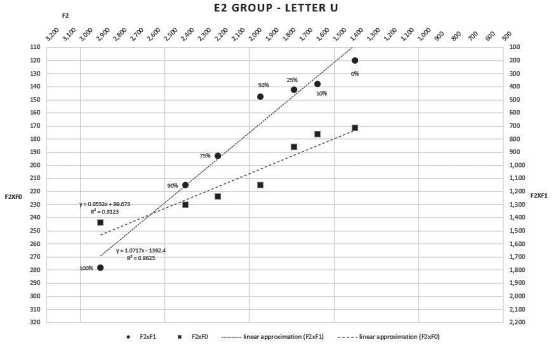

Fig 124. Letter 'U' – Group E2

the group E1, (0.0532, 99.673, 0.9123) for the group E2, (0.0511, 120.42, 0.9637) for the group UK, and (0.0508, 102.09, 0.9894) for the group US. Checking for the adequacy of the linear models to describe the percentiles, the r-squared parameters yielded high correlation with their values being greater than 0.9. As for the slopes, S1 was relatively higher than the group of natives and the other groups of students had values close to the number scored by the group US.

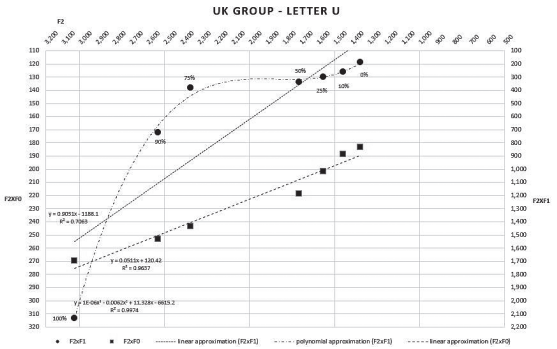

Fig 125. Letter 'U' – Group UK

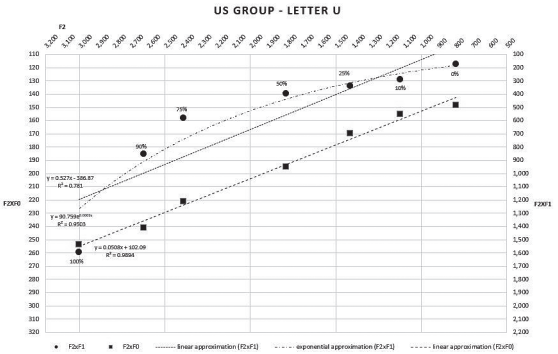

Fig 126. Letter 'U' – Group US

3.22 Letter 'V'

This letter tend to be mixed up with the letter 'B' by the students, so that some similarities in the temporal analysis were expected between the two letters. The results are shown in figures 127 through 132. In fact, the percentiles of the formants for the group JP were approximated by the curve defined by the triplet given by (1.1288, -1535, 0.956), whereas the group S1 was established

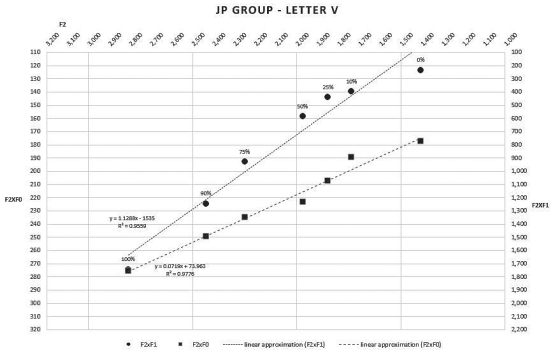

Fig 127. Letter 'V' – Group JP

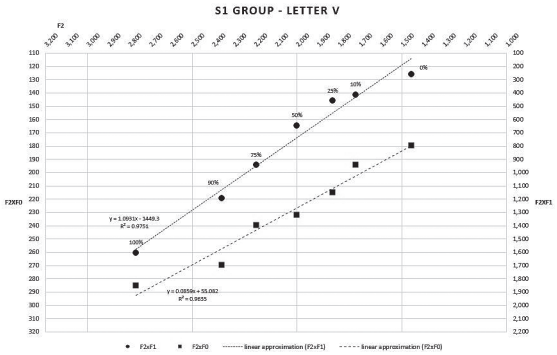

Fig 128. Letter 'V' – Group S1

by (1.0931, -1449.3, 0.9751), the group E1 by (1.0919, -1453, 0.9368), the group E2 by (1.1908, -1688.9, 0.9411), the group UK by (1.0979, -1802.2, 0.7582), and the group US by (0.7028, -896.69, 0.7421). Since the r-squared values of the groups of students were relatively large in the sense that they scored higher than 0.9, the behavior of the percentiles were pretty well described by the linear models. On the contrary, the groups of native speakers had nonlinearly varying percentiles.

Now, the percentiles of the pitch frequencies were (0.0719, 73.963, 0.9776) for the group JP, (0.0859, 55.082, 0.9635), for the group S1, (0.0586, 104.57, 0.9033) for the group E1, (0.0702, 62.515, 0.9319), for the group E2, (0.0738, 47.66, 0.9451) for the group UK, and (0.0659, 54.131, 0.9806) for the group UK. From these it is clear that the linear models fitted well on all groups with relatively high degree of fitness.

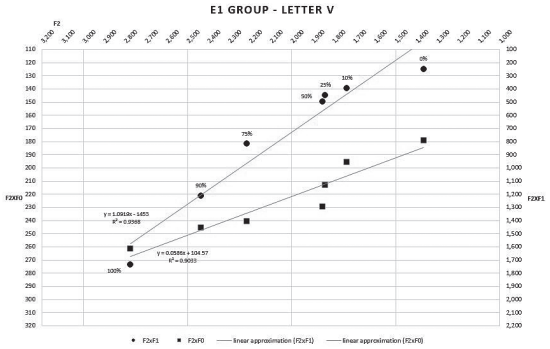

Fig 129. Letter 'V' – Group E1

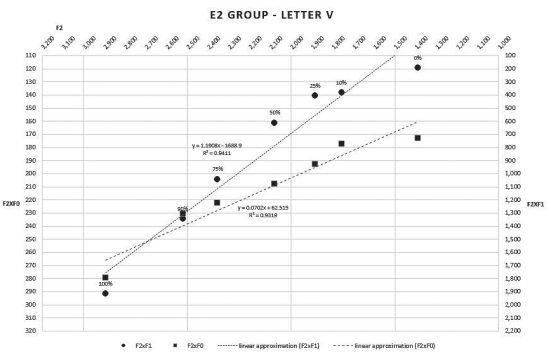

Fig 130. Letter 'V' – Group E2

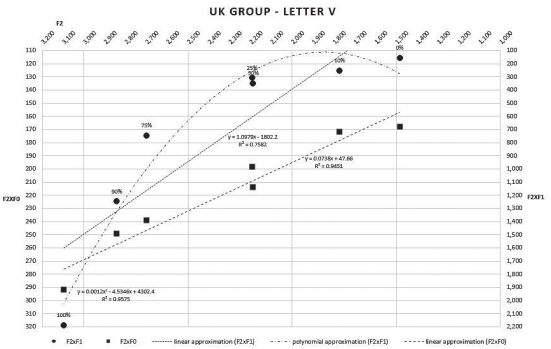

Fig 131. Letter 'V' – Group UK

Fig 132. Letter 'V' – Group US

3.23 Letter ‘W’

Figure 133 shows that the percentiles of the formants for the group JP had an approximant defined by the triplet (1.0539, -1289.7, 0.9802) whereas figure 134 through 138 display the computation results for the remaining groups. In fact, the group S1 read (1.1185, -1422.2, 0.9839), E1 measured (0.9682, -1118.2, 0.9512), E2 scored (1.0694, -1319.3, 0.9869), UK marked (0.7816, -780.13, 0.8859), and US gave (0.5574, -334.86, 0.8984). Since the values of the r-squared terms scored higher than 0.85, these linear models are reasonably acceptable. Furthermore, the slopes of the groups of the students were higher than the groups of natives.

For the analyses of the pitch frequency percentiles, the results show that the group JP was approximated by the triplet (0.0753, 65.044, 0.9581) whereas the group S1 gave (0.096, 28.92, 0.981), the group E1 output (0.0655, 95.337, 0.9424), the group E2 generated

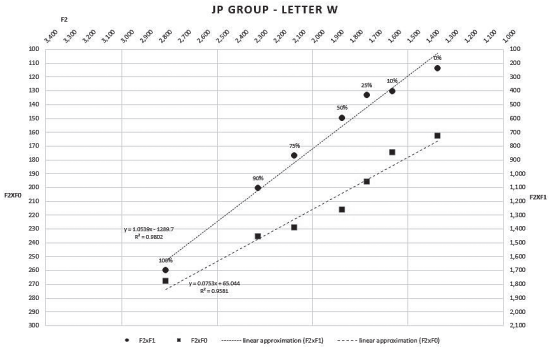

Fig 133. Letter ‘W’ – Group JP

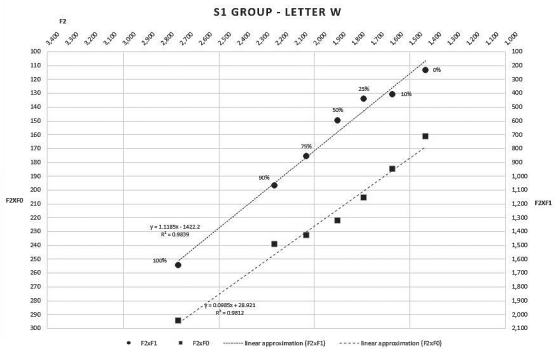

Fig 134. Letter ‘W’ – Group S1

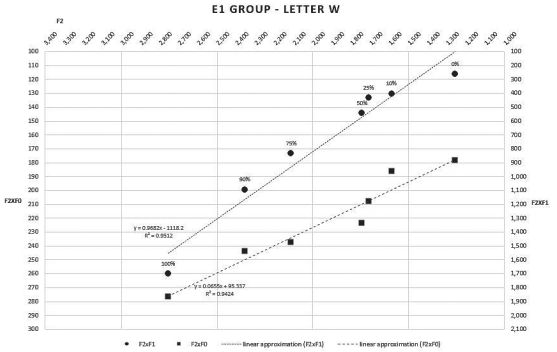

Fig 135. Letter ‘W’ – Group E1

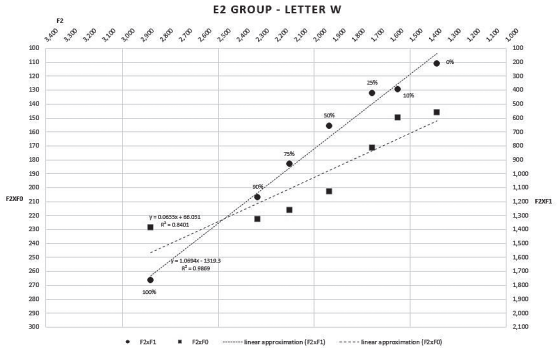

Fig 136. Letter ‘W’ – Group E2

(0.0633, 66.031, 0.8401), the group UK produced (0.0513, 111.43, 0.9698), and the group US rendered (0.062, 93.978, 0.9884). Apart from the group E2 that had r-squared value at 0.84, the other groups had values greater 0.9 which means that the linear models express the behavior of the percentiles. On the other hand, the slopes suggest that the groups of students had in general variation ratios more related to US and lesser to UK.

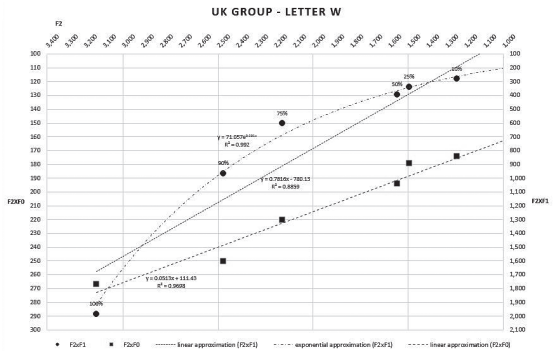

Fig 137. Letter 'W' – Group UK

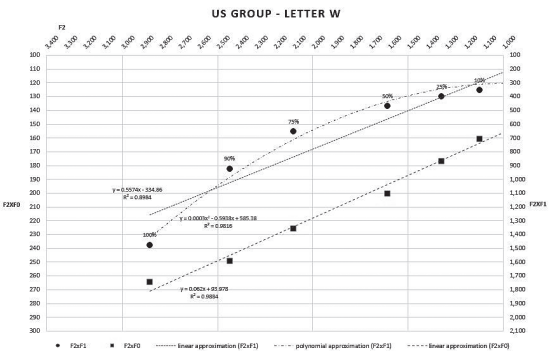

Fig 138. Letter 'W' – Group US

3.24 Letter 'X'

The results are shown in figures 139 through 144. As seen in the graphs, the percentiles of the formants were approximated by linear curves, which are characterized by the triplets as following: (1.1439, -1318.5, 0.9671) for the group of JP, (1.1776, -1425.9, 0.9591) for the group S1, (1.0203, -1036.9, 0.9691) for the group E1, (1.2609, -1552.2, 0.9657) for the group E2, (0.9752, -953.47, 0.9655) for the group UK, and

Fig 139. Letter 'X' – Group JP

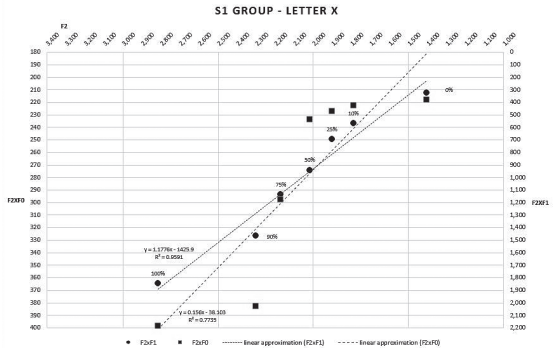

Fig 140. Letter 'X' – Group S1

(0.8775, -798.93, 0.9751) for the group US. The values of the r-squared terms, which were all greater than 0.95, assure that the linear models in fact describe very well the distribution of the percentiles. As for the analysis of the slopes, the triplets show that the students had higher values than the native speakers, which means that the students opened/ rounded the lips wider than the natives.

As far as the percentiles of the pitch frequencies are concerned, the triplet representing the linear approximants for the group JP was (0.1041, 42.362, 0.816), for the group S1 it was (0.156, -38.103, 0.7735), for the group E1 the trio read (0.0797, 99.932, 0.8025), for the group E2 (0.0697, 73.265, 0.8945), for the group UK (0.048, 120.08, 0.9711), and for US (0.0468, 123.91, 0.8935). Note that the slopes of the students were higher than those of the group of natives.

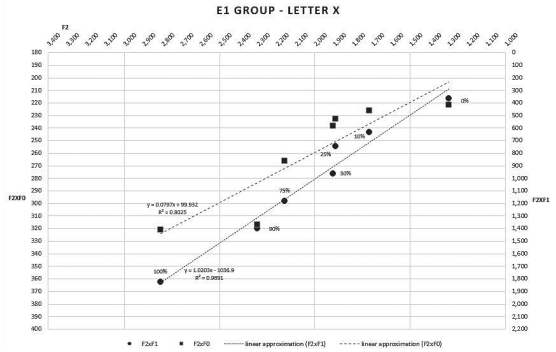

Fig 141. Letter 'X' – Group E1

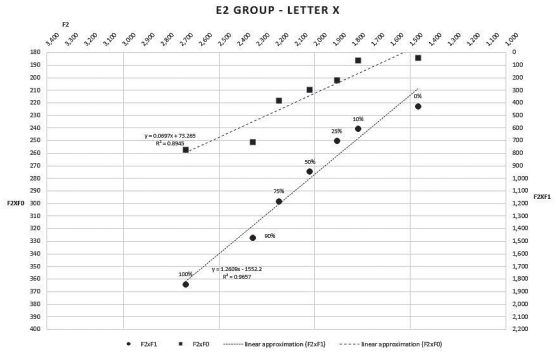

Fig 142. Letter 'X' – Group E2

Fig 143. Letter 'X' – Group UK

Fig 144. Letter 'X' – Group US

3.25 Letter 'Y'

Since some people read the letter 'Z' as 'zi:' and others as 'zed', it was excluded from the analyses. So, the letter 'y' is the last letter to be considered in this work. The results are shown in figures 145 through 150 and they tell us that the percentiles of the formants for the group JP were expressed by the triplet (0.7207, -466.9, 0.975) whereas for the group S1 the values were (0.7782, -602.13, 0.9786), for the group E1 the numbers were (0.6905, -437.13, 0.958), for the group E2 the scores were (0.6794, -331.99, 0.9577), and for the groups UK and US the values were respectively (0.6607, -288.6, 0.8769) and (0.524, -140.84, 0.9759). The r-squared parameters show that the linear models explain fairly well the distribution of the percentiles. Yet, the slopes of the groups of students tended to have values near the group US.

Focusing on the percentiles of the pitch frequencies, the triplet for the group JP was (0.066, 92.043, 0.9524), for

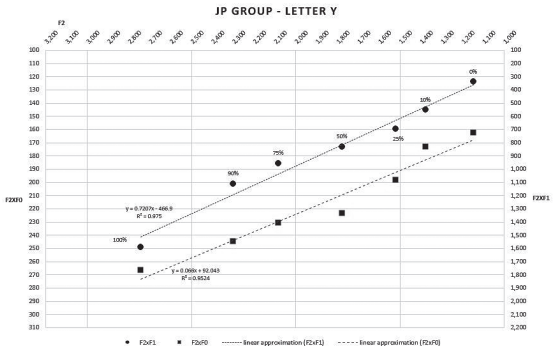

Fig 145. Letter 'Y' – Group JP

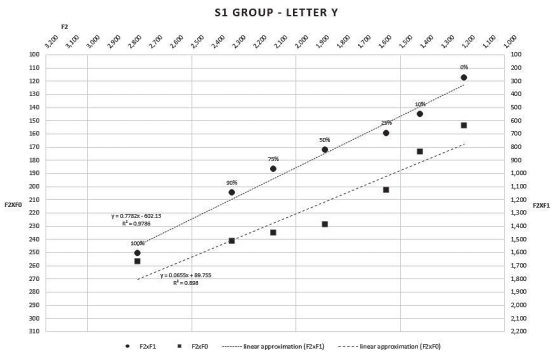

Fig 146. Letter 'Y' – Group S1

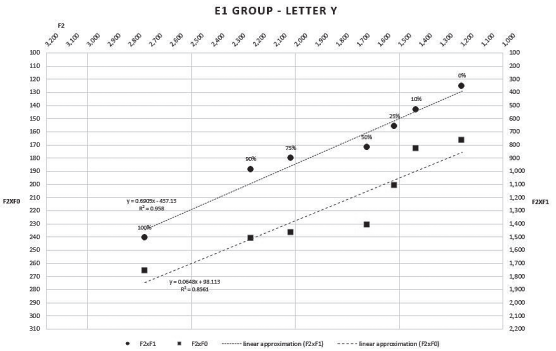

Fig 147. Letter 'Y' – Group E1

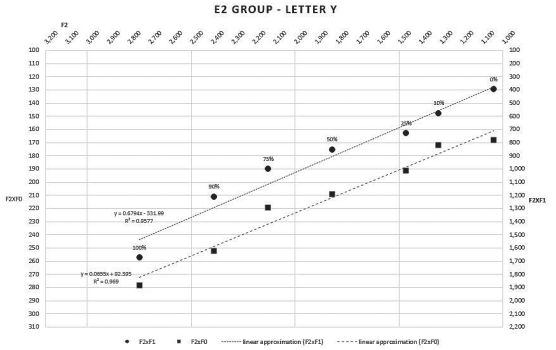

Fig 148. Letter 'Y' – Group E2

the group S1 (0.0655, 89.755, 0.898), for the group E1 (0.0648, 98.113, 0.8561), for the group E2 (0.0655, 92.395, 0.969), for the group UK (0.0205, 160.45, 0.9655), and for the group US it read (0.0511, 100.62, 0.9939). Similar to the percentiles of the formants, the behavior of the pitch frequencies were described by the linear models, and the slopes of the groups of students were closer to the group US than UK.

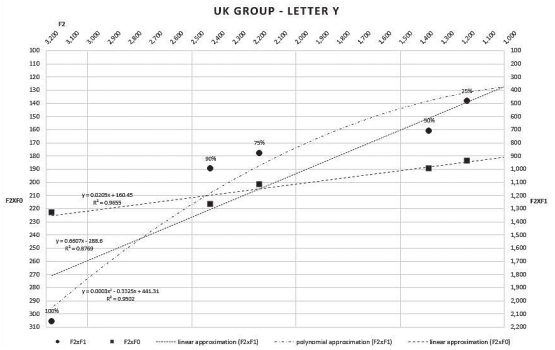

Fig 149. Letter 'Y' – Group UK

Fig 150. Letter 'Y' – Group US

4 DISCUSSIONS

Tables 1 and 2 summarize the results so far. Rating the linear model according to the values of r-squared parameters as very strong (VS) for values greater than or equal to 0.9 - not including the right bound numbers - and defining fairly strong (FS) for values between 0.8 and 0.9; strong (S) for the interval from 0.7 and 0.8; not strong (NS) for 0.6 to 0.7; and inadequate (I) for smaller values.

Table 1 shows that the groups of students were in general "VS" whereas the groups of the native speakers varied a lot. Thus, as far as the linearity is concerned, the groups of students differed relatively a lot from those of natives for the categories of the alphabet [i:/i] and [u:] in the way the percentiles of the formants varied along the time line.

For the percentiles of the pitch frequencies shown in Table 2, both the students and natives had values well expressed by the linear model so that no conspicuous difference was noted. Note, however, that only a statistical testing will give useful information to whether they are similar or not.

Table 1 Fitness of the linear model to express the formant percentiles

VS: very strong, FS: fairly strong, S: strong, NS: not strong; I:inadequate							
Category	**Letter**	**JP**	**S1**	**E1**	**E2**	**UK**	**US**
[i:/i]	B	VS	FS	VS	FS	I	NS
	C	VS	VS	VS	VS	FS	FS
	D	VS	FS	VS	VS	NS	NS
	E	FS	FS	VS	FS	I	NS
	G	VS	FS	VS	FS	S	NS
	P	VS	VS	VS	FS	NS	S
	T	VS	VS	VS	VS	NS	S
	V	VS	VS	VS	VS	S	S
	Z	-	-	-	-	-	-
[e]	F	VS	VS	VS	VS	VS	VS
	L	VS	VS	VS	VS	VS	VS
	M	VS	VS	VS	VS	VS	VS
	N	VS	VS	VS	VS	VS	FS
	S	VS	VS	VS	VS	VS	VS
	X	VS	VS	VS	VS	VS	VS
[ei]	A	VS	FS	VS	VS	NS	FS
	H	VS	VS	VS	VS	VS	FS
	J	VS	VS	VS	VS	FS	FS
	K	FS	FS	VS	FS	S	FS
[u:]	Q	VS	VS	VS	VS	VS	FS
	U	VS	VS	VS	VS	S	S
	W	VS	VS	VS	VS	FS	FS
[ai], [o] [r]	I	VS	VS	VS	VS	FS	VS
	Y	VS	VS	VS	VS	FS	VS
	O	VS	VS	VS	VS	VS	VS
	R	VS	VS	FS	VS	VS	VS

Table 2 Fitness of the linear model to express the pitch percentiles

VS: very strong, FS: fairly strong, S: strong, NS: not strong; I:inadequate							
Category	**Letter**	**JP**	**S1**	**E1**	**E2**	**UK**	**US**
[i:/i]	B	FS	VS	S	FS	FS	VS
	C	VS	VS	VS	VS	S	VS
	D	VS	VS	VS	VS	VS	FS
	E	VS	VS	FS	VS	NS	VS
	G	VS	VS	S	VS	FS	FS
	P	VS	VS	VS	FS	S	FS
	T	VS	VS	VS	VS	S	FS
	V	VS	VS	VS	VS	VS	VS
	Z	-	-	-	-	-	-
[e]	F	S	S	S	S	FS	FS
	L	VS	VS	VS	VS	VS	VS
	M	VS	VS	VS	FS	VS	VS
	N	FS	FS	VS	FS	FS	VS
	S	FS	S	S	S	VS	FS
	X	FS	S	FS	FS	VS	FS
[ei]	A	VS	VS	VS	VS	NS	FS
	H	VS	FS	VS	FS	VS	VS
	J	VS	VS	FS	VS	VS	VS
	K	VS	VS	VS	FS	S	VS
[u:]	Q	VS	FS	FS	VS	VS	VS
	U	VS	VS	FS	VS	VS	VS
	W	VS	VS	VS	FS	VS	VS
[ai], [o] [r]	I	FS	VS	FS	VS	VS	VS
	Y	VS	FS	FS	VS	VS	VS
	O	FS	FS	FS	VS	FS	VS
	R	S	S	FS	FS	VS	FS

Acknowledgement

The authors would like to thank the students of Yonezawa Women's College who provided us with the voice signals and helped us with the experiments. Thanks also go to the staff members, colleagues and friends for the support and cooperation.

REFERENCES

Barreda, S. and Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. J. Acoust. Soc. Am., 131(1), pp.466-477.

Binns C. and Culling J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. J. Acoust. Soc. Am., 122(3), pp.1765–1776.

Carrol J., Tiaden S., and Zeng F. G. (2012). Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. J. Acoust. Soc. Am., 130(4), pp.2054-2062.

Fujisaki, H., and Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels, IEEE Trans. Audio Electroacoust., AU-16, pp. 73–77.

Hillenbrand, J. M. and Clark M.J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. Attention, Perception, & Psychophysics, 71 (5), pp.1150-1166.

Izuta, G. (2013). An assessment of the utterances of some English vowels spoken by Japanese Students – A survey at a junior college for women in Japan , Bulletin of Yonezawa Women's Junior College, Vol.49, p.63-82.

Izuta, G. (2014a). Does the production mechanism of the typical Japanese utterances affect the English Speech? Some insights from a phonetic analysis perspective, Reports of The Institute for Culture in Life of Yonezawa Women's Junior College, Vol. 41, pp.39-57.

Izuta, G. (2014b). Towards understanding how young Japanese female college students pronounce the letters of English alphabet - Part I: general analysis of the formants F1 and F2, Bulletin of Yonezawa Women's Junior College, Vol.50, p.59-87.

Izuta, G. (2015a). Towards understanding how young Japanese female college students pronounce the letters of English alphabet - Part II: analysis of the pitch frequency modulation, Reports of The Institute for Culture in Life of Yonezawa Women's Junior

College, Vol. 42, pp.17-74.

Izuta, G. (2015b). On How Young Japanese Female College Students Say the English Alphabet: A Comparison of the Fundamental Frequencies F0s with Native Speakers of English Language, Proceedings of The Asian Conference on Second Language Acquisition and Teacher Education, , SLATE 2015/8/2-4 Hiroshima, digital format, ISSN: 2189-2164, 6 pages.

Izuta, G. (2015c). A Study on How Young Japanese Female College Students Say the English Alphabet: Comparison of the Fundamental Frequencies F0s across the Sub-Groups and Along Different Percentiles, Proceedings of The Asian Conference on Second Language Acquisition and Teacher Education, SLATE 2015/8/2-4 Hiroshima, digital format, ISSN: 2189-2164, 8 pages.

Izuta, G. (2015d). Formant Analysis of the English Alphabet Read by Japanese Female Students and Native Speakers, Proceeding: 1st International Conference on Teaching & Learning (ICTL 2015), Langkawi, Malaysia 14-15 September 2015, ISBN 978-967-13637-1-3, pp.18-22

Izuta, G. (2015e). A Study of the English Sounds Made by Japanese Female Students Based on the Percentile Analysis, Proceeding: 1st International Conference on Teaching & Learning (ICTL 2015), Langkawi, Malaysia 14-15 September 2015, ISBN 978-967-13637-1-3, pp.23-27.

Richards, J. C. and Rogers, S. T. (2001), Approaches and Methods in Language Teaching, Cambridge University Press.

Ladefoged, P. (1993). “A Course in Phonetics”. Harcourt Brace Jovanovich, Orlando.

Ladefoged, P. (2007). “Phonetic Data Analysis”. Blackwell Publishing, Carlton.

Moore, A. (2001). The effects of formal instruction on EFL pronunciation acquisition: a case study from Germany, MA dissertation, University of Surrey.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages, TESOL quarterly, XXV (3), pp.481-520.

Ohata, K. (2004). Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners, Asian EFL Journal, 6 (4), pp. 1-19.

Patterson, R. D. and Johnsrude, I. S. (2012). Functional imaging of the auditory processing applied to speech sounds. Phil. Trans. R. Soc., B-363, pp. 1023–1035.

Smith, B. (2012). Pronunciation Patterns of Japanese Learners and their Implications or Teaching, Polyglossia, Volume 23, October 2012, pp. 199-206.

Sugimoto, M. (2014). Word Accent in Japanese and English - What Are the Differences? Hitsuji Shobou, Japan.

Riney, T. and Anderson-Hsieh, J. (1993). Japanese pronunciation of English. Vol. 15, No.1, pp. 21-36.

Titze, R. (1994). I. Principles of Voice Production. Prentice-Hall, Englewood Cliffs, NJ.

Zatorre, R. J. and Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. Phil. Trans. R. Soc., B-363, 1087–1104.